Analisi e misura di segnali comportamentali acquisiti su utenti anziani tramite sensori domotici.

Analysis and measurement of behavioral signals acquired on elderly users through home automation sensors.

Relatore : **Prof. Ing. Gian Marco Revel** Laureando : **Domenico Rosica**

Correlatore: **Ph.D. Sara Casaccia**

Indice

1. **Introduzione**
   1.1 Obiettivo della tesi

2. **Stato dell'arte**
   2.1 Sistemi domotici per Active and Assisted Living

3. **Progetto H@H**
   3.1 Descrizione del progetto
   3.2 Obiettivi del progetto
   3.3 Metodi di acquisizione

4. **Materiali e metodi**
   4.1 Configurazione sensori PIR e luci installati negli appartamenti
   4.2 Dati e segnali acquisiti
   4.3 Analisi dei dati
   4.4 Correlazioni attraverso Pearson e Spearman
   4.5 Metodi di Machine Learning
   4.6 Metodo di valutazione: **Cross-Validation**
   4.7 Metodo di valutazione: **Hold-out Validation**

5. **Risultati**
   5.1 Risultati delle correlazioni
   5.2 Risultati del ML

6. **Conclusioni**

Ringraziamenti

Bibliografia

1. Introduzione

Nell'ultimo decennio si è sempre più diffuso il concetto di Domotica ovvero la scienza interdisciplinare che si occupa dello studio delle tecnologie atte a migliorare la qualità della vita nella casa [1].

Questo tipo di impianti viene utilizzato anche in altri ambiti quali: ospedali, case per persone disabili o abitazioni di persone anziane, allo scopo di fornire, a chi vi è all'interno, degli strumenti di sostegno; si parla in questo caso di "domotica assistenziale", la quale si suddivide in due tipologie: la domotica per l'autonomia e la domotica per la sicurezza [1].

La prima prevede automazioni di diverso tipo che, integrate da un sistema di gestione unitario, permettono anche a chi ha difficoltà, soprattutto motorie, di avere un controllo autonomo della casa, mentre la seconda, pensata soprattutto per l'anziano o per la persona con disabilità mentali, basa il suo principio di funzionamento su apparecchiature indossabili o posizionati all'interno della casa e il suo scopo è quello di garantire condizioni di benessere e sicurezza [1].

In ambito internazionale invece si parla di Active and Assisted Living (AAL) che indica la disciplina che si occupa di migliorare la qualità della vita degli anziani e dei disabili investendo sulle nuove tecnologie [2].

Ad oggi sono sempre più numerosi i paesi che si servono della tecnologia per l'assistenza nella vita di tutti i giorni. Quest'ultima offre nuove opportunità per migliorare il benessere e la qualità della vita dei suoi utenti e sta diventando una necessità soprattutto nel campo dell'assistenza sanitaria. Secondo l'Organizzazione Mondiale della Sanità (OMS), la popolazione anziana mondiale è in aumento con il conseguente aumento delle risorse sanitarie e mediche e una maggiore dipendenza da badanti, famiglie e società [2].

Per ovviare a questo trend, molta attenzione, negli ultimi decenni, è stata data alle cosiddette "Smart Homes", specialmente nella comunità di ricerca che ha preso in esame i vantaggi di un monitoraggio e di un'automazione potenziati servendosi di sensori e attuatori per fornire assistenza. La combinazione di assistenza sanitaria e Smart Homes dà origine al termine "Smart Home Healthcare", in questo modo oltre a una efficienza dal punto di vista del trattamento si potenzia molto la prevenzione [2].

Si può quindi definire una "Smart Home" una casa o abitazione equipaggiata di sensori e dispositivi collegati in rete che aumentano la funzionalità della casa aggiungendo intelligenza, automazione, controllo, consapevolezza contestuale, adattabilità e funzionalità sia da remoto che a livello locale, al fine di migliorare la salute e il benessere degli occupanti e l'assistenza nella fornitura di servizi sanitari [1]. In riferimento a quanto spiegato sopra, in questa tesi si dimostrerà come dispositivi domotici installati in casa di anziani possano fornire delle informazioni utili per lo studio e l'analisi dello stile di vita quotidiano degli individui.

1.1 Obiettivo della tesi

In riferimento a quanto spiegato sopra, in questa tesi si dimostrerà come i dispositivi domotici installati in casa di anziani possano fornire delle informazioni utili per lo studio e l'analisi dello stile di vita quotidiano degli individui. Lo scopo sarà quello di determinare con buona accuratezza lo stato fisico e mentale degli anziani partendo dai dati raccolti dai sensori. In questa maniera si è in grado di monitorare la vita quotidiana degli anziani fornendo loro l'assistenza necessaria qualora ci fosse bisogno. Per questo tipo di studio verranno presentate delle procedure di analisi nell'ambiente di sviluppo Matlab, facendo particolare riferimento alle tecniche di Machine Learning utilizzate.

2. Stato dell'arte

Le tecnologie installate nelle case utilizzate per fornire assistenza in casa ai pazienti con problemi deambulatori, agli anziani e persone con disabilità sono indicate con Smart Home Healthcare. Ad oggi, oltre che assistere persone con deficit mentale o fisico, si tende a soddisfare anche gli individui sani, monitorando la salute e il benessere degli occupanti utilizzando sensori indossabili o incorporati nell'ambiente circostante.

Dato che lo sviluppo della sanità e della medicina continuano a tradursi in un tasso di mortalità inferiore, il ramo delle risorse mediche tradizionali impone l'esigenza di soluzioni diverse, quali le case intelligenti, che possono rivelarsi efficaci in termini di costi per aiutare anziani e disabili a rimanere a casa più a lungo in modo non invadente, consentendo a quest'ultimi una maggiore indipendenza e qualità di vita, riducendo la possibilità di isolamento sociale, grazie all'utilizzo di componenti domestici altamente tecnologici quali i sensori, gli azionatori, i regolatori e le varie apparecchiature integrate con i dispositivi biomedicali, i quali possono creare 7 giorni alla settimana, H24, soluzioni sanitarie per i più vulnerabili [1].

Figura 1: Smart Home

Recenti studi evidenziano il progressivo invecchiamento della popolazione nel panorama del 21° secolo; gli esperti prevedono una ridistribuzione demografica senza precedenti, in cui entro il 2050 la proporzione di anziani tenderà a raddoppiare, passando dall'11% al 21% della popolazione totale. Nei prossimi 5 anni, per la prima volta nella storia dell'umanità, il numero di individui di età uguale o superiore a 65 anni supererà quello dei bambini al di sotto dei 5 anni e il segmento degli ultraottantenni, entro il 2050, risulterà praticamente quadruplicato (figura 2) [3].

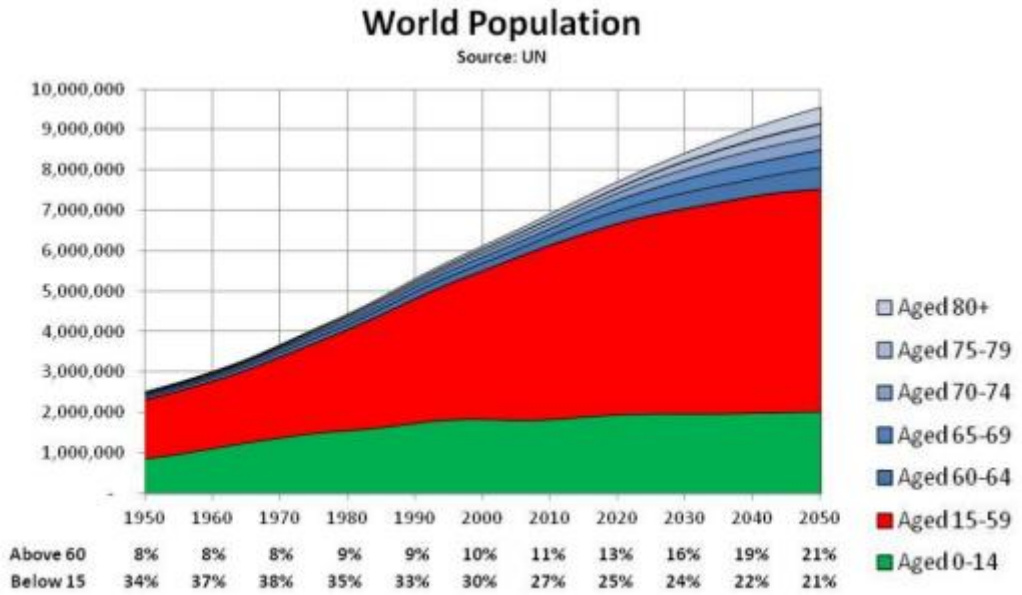

Figura 2: Crescita mondiale della popolazione (Nazioni Unite)

Questo trend sta spostando lo sviluppo tecnologico verso nuove frontiere in quanto la domanda per sistemi innovativi per l'assistenza medica domestica sta crescendo molto velocemente con il passare degli anni, e poiché persone diverse hanno esigenze diverse, la fornitura di assistenza deve essere adattata a ciascuna richiesta. Molte aziende e centri di ricerca, in tutto il mondo, si sono specializzate in questo settore per fornire questi servizi: un'applicazione italiana del concetto di "Smart Home" è il progetto "Casattenta" sviluppato da T3Lab [4].

Le tecnologie per le Smart Homes sono sempre più in vendita, tanto che il mercato globale per le apparecchiature intelligenti è proiettato a crescere da 40 milioni nel 2012 a 26 miliardi di dollari nel 2019 [5]. La crescita del mercato dipenderà in ultima analisi dai potenziali utenti che percepiranno chiaramente i benefici di queste nuove tecnologie; il più significativo ostacolo all'adozione delle case intelligenti è il costo iniziale, seguito dalla mancanza di consapevolezza sul relativo risparmio energetico e problemi legati alla privacy [6].

In particolare, per questa tesi, sono stati presi come riferimento due articoli:

I. Unsupervised Machine Learning for Developing Personalised Behaviour Models Using Activity Data
II. A Plug and Play IoT Wi-Fi Smart Home System for Human Monitoring

Il primo articolo ha come obiettivo quello di proporre un metodo per descrivere i modelli comportamentali dei singoli utenti a partire da un'analisi dei dati di un minimo numero di sensori e un approccio "cieco" per il riconoscimento delle attività. Ovviamente vi sono alcune differenze nel metodo seguito ovvero non vengono raccolti dati dalle accensioni degli interruttori delle luci, viene adottato un approccio "cieco" ovvero non si conoscono le effettive condizioni del soggetto ovvero non dispongono di questionari, vi è una suddivisione dei dati in fasce giornaliere (mattina, tarda mattina, pomeriggio…etc) e si pratica un confronto dei modelli di comportamento quotidiani in stanze diverse [7].

Il secondo articolo ha come obiettivo quello di rilevare tramite l'utilizzo di sensori domotici cambiamenti di routine dei soggetti. Ovviamente vi sono alcune differenze nel metodo seguito ovvero l'utilizzo di un grande numero di sensori (sensori di temperatura, sensori magnetici su porte e finestre, più di 10 sensori PIR per abitazione), confronto dei modelli di comportamento quotidiani in stanze diverse e lo studio viene fatto su pazienti con particolari malattie degenerative (Parkinson, Alzheimer). Vi sono comunque molte similitudini quali compatibilità strumentale (utilizzo di sensori PIR), ricerca di possibili correlazioni mediante indici di correlazione, conoscenza dal principio dello stato fisico e mentale dei soggetti [8].

2.1. Sistemi domotici per utenti anziani

Negli ultimi anni, Internet-of-Things (IoT) ha suscitato molta attenzione da parte di ricercatori, imprenditori, e giganti tech in tutto il mondo. L'IoT è una tecnologia emergente che ci permette di connettere una grande varietà di dispositivi e sistemi quotidiani come sensori, attuatori, elettrodomestici, computer e telefoni cellulari, in grado di comunicare con altri dispositivi e con gli esseri umani [9].

La Smart Homes, esempio di e-health integrata e tecnologia di vita assistita, è un'applicazione dell'IoT che è in grado di svolgere un ruolo cruciale nel cambiamento in che può potenzialmente ricoprire un ruolo fondamentale nella rivoluzione del sistema sanitario per gli anziani. In una casa intelligente, sensori e attuatori sono collegati attraverso una Personal Area Network (Pan) o Wireless Sensor Network (WSN) [9].

Sensori biomedici indossabili come l'ECG, l'EMG, l'EEG e quelli relativi alla misurazione della temperatura corporea e della saturazione dell'ossigeno possono essere collegati in una Wireless Body Area Network (WBAN) o Body Sensor Network (BSN) per ottenere misurazioni automatiche, continue e in tempo reale di segnali fisiologici [9].

Figura 3: Wireless Body Area Network con piattaforma di assistenza sanitaria connessa

Le tecnologie per il benessere fisico e mentale hanno caratteristiche eterogenee; queste tecnologie sono classificate nelle seguenti principali aree tecnologiche:

- sistemi ambientali di monitoraggio, come telecamere, sensori laser, sensori di movimento, sensori letto; quest'ultimi consentono di analizzare e di monitorare la routine dei soggetti, con lo scopo di evidenziare il verificarsi di possibili anomalie nelle loro abitudini e disturbi del sonno. Queste alterazioni sono spesso indispensabili per diagnosticare precocemente le mattie come il deficit cognitivo [11]. Nella figura 4 è possibile capire il principio di funzionamento di un sensore PIR. Tale sensore si attiva quando rileva il movimento di una fonte di calore dando in output un numero non nullo.

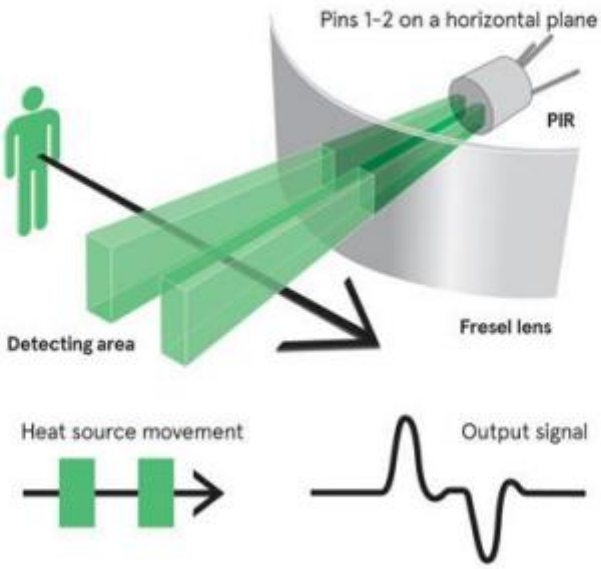

Figura 4: funzionamento di un sensore di movimento ad infrarossi (PIR sensor, Passive Infrared Sensor)

- sistemi multisensoriali indossabili utilizzati per monitorare e misurare segnali fisiologici, ad esempio cinture toraciche e braccialetti, che permettono il controllo delle attività quotidiane e le misure dei segnali fisiologici. Gli strumenti più utilizzati sono: accelerometri e giroscopi a tre assi, sensori per il rilevamento della temperatura corporea e per l'ossigenazione sanguigna Per poter monitorare al meglio i parametri vitali, è consigliabile un'interazione tra i due sistemi (ad esempio, ambientale e indossabile), con un canale di comunicazione dedicato [11].

Figura 5: moderno smartwatch con rilevamento ECG e saturimetro digitale

- robot intelligenti per la cura domestica e il monitoraggio dell'attività, sostenendo il benessere fisico e psicologico degli anziani. Forniscono un due tipi di assistenza: passiva e attiva.

L'assistenza passiva oltre monitorare gli utenti, li aiuta in semplici azioni come ad esempio ritrovando oggetti persi in precedenza, aiutandoli nella ricerca di oggetti smarriti e assumono un ruolo di segreteria per ricordare gli impegni del giorno. Al contrario, l'assistenza attiva riguarda la capacità del robot di interagire autonomamente con gli anziani proponendo una serie di attività che possono stimolare sia la loro mente che il corpo [11].

Figura 6: Zora Robot nell'assistenza per gli anziani

3. Progetto H@H

figura 7: HEALT@HOME

3.1 Generalità sul progetto

Health@Home è un progetto finanziato dal MIUR che si focalizza nell'assistenza domestica. H@H ha lo scopo di sviluppare una rete di sensori residenziali integrati, ovvero la combinazione di dispositivi domotici e biomedicali, per aiutare le persone anziane a migliorare il loro stile di vita all'interno dell'abitazione [12][13][14]. Il progetto iniziato nel 2014 e terminato nel 2019 prevedeva di testare i sistemi sviluppati in alcune sperimentazioni tra cui una nella regione Veneto. Tale sperimentazione è stata portata avanti dall'Università Politecnica delle Marche, Università La Sapienza, HomeLab e ha coinvolto tredici utenti distribuiti in otto abitazioni nel comune di Oderzo. La sperimentazione mira a fornire l'integrazione delle tecnologie dell'informazione e della comunicazione (TIC), per supportare gli individui attraverso le metodologie di vita attiva e assistita (AAL). Per la selezione degli appartamenti alcuni requisiti sono stati presi in considerazione: in primo luogo, il Consorzio di progetto ha individuato la A.T.E.R. (azienda territoriale per l'Edilizia residenziale) di Treviso come partner di progetto, pertanto le case sono state identificate nella regione Veneto dove opera l'azienda. Sono stati inseriti nel progetto la AULSS (azienda unica locale sanitaria sociale) e la OIC (Fondazione Opera Immacolata Concezione ONLUS). La città che più si presta al progetto è quella di Oderzo. Questa è risultata idonea per i requisiti e inoltre vi operano A.T.E.R., AULSS e OIC. Gli utenti sono stati assunti secondo alcune caratteristiche specifiche basate su: età, genere, patologie, gruppo familiare, stato sociale, condizioni psicologiche/fisiologiche [12].

3.2 Principali scopi del progetto

La struttura H@H è stata costruita per fornire servizi per gli utenti più anziani tramite una rete di sensori, installati nell'ambiente domestico, che raccoglie informazioni di varie tipologie, parameteri fisiologici e comportamentali, memorizzate nel cloud. Sono stati eseguiti dei test preliminari atti a comprovare la fattibilità dell'architettura, ovvero l'efficienza della rete dei sensori, la corretta memorizzazione delle informazioni nel cloud e la definizione dei servizi attraverso tecniche di analisi dei dati. Grazie l'applicazione H@H creata per la coordinazione dei dispositivi biomedicali e domotici, è possibile acquisire i parametri fisiologici, quali ECG, HR, temperatura timpanica, glicemia, pressione arteriosa, saturazione dell'ossigeno e peso corporeo. Attraverso l'applicazione, installata per gli utenti più anziani, è anche possibile gestire il sistema domotico. Tutti i dati nel cloud saranno utilizzati per definire servizi personalizzati per gli utenti [12].

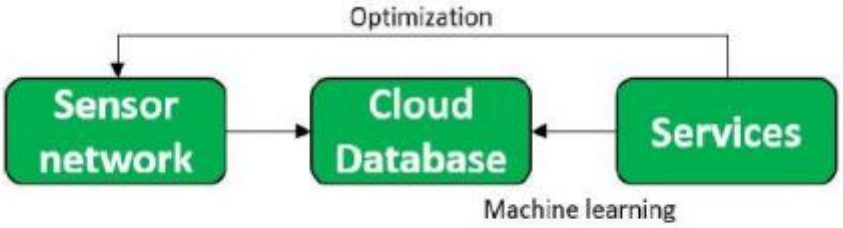

Figura 8: schema del flusso di dati proveniente dalla rete di sensori per i servizi sviluppati per l'utente

3.3 Metodi di acquisizione

DISPOSITIVI BIOMEDICALI	DISPOSITIVI DOMOTICI
Cintura multi-parametrica	Sensore Frigorifero smart di Whirpool
sfigmomanometro	Sensori luci
Pulsossimetro	Sensori PIR
Glucometro	
Termometro timpanico	
Bilancia	

La struttura del sistema residenziale integrato è composta da diverse apparecchiature domotiche e componenti intelligenti integrati nello stesso livello di scambio di dati all'interno della abitazione. Le componenti biomedicali sono caratterizzate da un utilizzo dinamico per facilitare l'utente durante l'acquisizione dei dati, monitorando i parametri fisiologici [12].

I dispositivi biomedicali sono:

• **Cintura multi-paramedica** per il monitoraggio della frequenza cardiaca, frequenza respiratoria, segnale accelerometrico, livello di attività, postura e prima derivazione del segnale elettrocardiografico (ECG) e respiratorio;

• **Sfigmomanometro** anch'esso utilizzato per la misura della frequenza cardiaca ma più in particolare per la misura della pressione sanguigna, diastolica e sistolica;

• **Pulsossimetro** per la misura della saturazione dell'ossigeno presente nel sangue e per verificare che la frequenza cardiaca acquisita sia stata misurata correttamente;

• **Glucometro** per la misura dei valori di glicemia sanguigna;

• **Termometro timpanico** per la misura della temperatura timpanica;

• **Bilancia** per la misura del peso corporeo.

L'Architettura software sviluppata per la rete dei sensori è stata divisa in due gruppi:

- il blocco "Service Bridge", che comprende tutti i dispositivi biomedicali, il frigorifero intelligente Whirlpool e il rilevatore della qualità dell'aria Elica (SNAP);

- il blocco "Home", il quale integra il sistema BTicino con il proprio protocollo di comunicazione e invia i dati direttamente al cloud per un'ulteriore analisi e sviluppo dei servizi. L'applicazione BTicino permette la misurazione e il controllo della temperatura ambientale, degli eventi comportamentali, dello stato di luce e dei consumi elettrici, raccogliendo tutti i cambiamenti di stato interni ed esterni e i messaggi di errore per ogni dispositivo connesso, avendo cura anche del processo di scoperta di ogni oggetto intelligente [12].

Figura 9: Architettura della rete di sensori H@H

4. Materiali e metodi

4.1 Configurazione sensori PIR e luci installati negli appartamenti

Ai fini dello studio comportamentale sulle abitudini di vita dei soggetti, i sensori PIR si rivelano molto utili ed efficaci. Il PIR, acronimo inglese di Passive InfraRed, è un sensore elettronico che misura i raggi infrarossi raccolti nel suo campo di vista, i quali emettono una radiazione luminosa propria. Permette di analizzare, per i corpi che hanno una temperatura diversa dallo zero assoluto (0 gradi Kelvin), uno spettro di frequenze non visibili all'occhio umano. Il termine "passivo", fa riferimento al fatto che questi dispositivi non emettono alcuna forma di energia ma lavorano esclusivamente rilevando l'energia, sotto forma di radiazioni, emessa dagli oggetti circostanti. I sensori PIR sono dunque alla base della costruzione di sensori di movimento.

Figura 10: esempio di PIR aventi diversa morfologia

In pratica, un PIR non rileva autonomamente un movimento ma repentine variazioni di temperatura che ne modificano lo stato iniziale. Quando qualcosa o qualcuno passa di fronte ad uno sfondo, come ad esempio un muro, precedentemente "fotografato" dal PIR come stato normale, la temperatura in quel punto si innalza istantaneamente, passando da quella della stanza a quella del corpo. Questo rapido cambiamento attiva il rilevamento. Lo spostamento di oggetti di temperatura identica rispetto allo sfondo, naturalmente, non viene rilevata. Tali dispositivi inoltre, al loro interno, posso essere costituiti da più sensori combinati tra loro in modo che, con le appropriate aggiunte elettroniche, si possa ricavare, dal rilevamento sequenziale dei sensori, la direzione del movimento. Questo è un modo efficace per evitare di rilevare “falsi allarmi”, così da aumentarne l’affidabilità e, più in generale, per raccogliere informazioni sulle varie direzioni di movimento all’interno di un sistema per il monitoraggio [15].

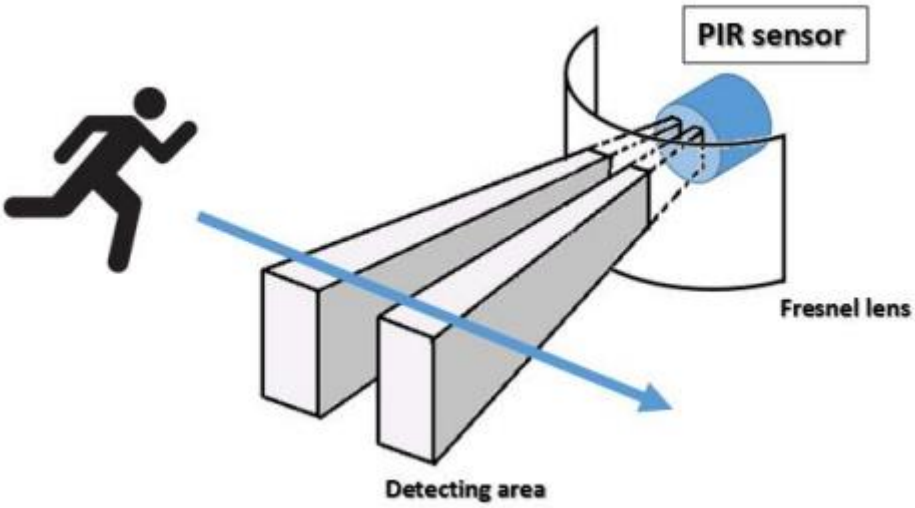

Figura 11: Funzionamento sensore PIR

È importante scegliere il posizionamento dei sensori per evitare la rilevazione di “falsi allarmi” all’interno delle “case pilota” del progetto H@H, alla luce di ciò i sensori sono stati collocati in modo tale da non essere in grado di recepire segnali al di fuori del campo visivo stabilito per il monitoraggio, ovvero non si vuole ad esempio che il PIR posizionato in camera rilevi movimenti in cucina. In questo modo la raccolta dati e quindi l’analisi stessa risulta essere vera, ciò rende l’analisi dei dati più fedele ed attendibile. In particolare i sensori sono stati installati in tutte le abitazioni nella camera da letto, nel soggiorno e nel bagno, ad eccezione delle case 3, 4,7 e 10 dove mancano rispettivamente il PIR in bagno nelle prime due, in camera nella 7. Di seguito, in figura 12, viene riportata la planimetria dell’abitazione 5 dove sono evidenziati i sensori PIR e il loro posizionamento.

- PIR 1 installato nella stanza Cucina/Soggiorno;

- PIR 2 installato nella Camera;

- PIR 3 installato nel Bagno.

Figura 12: Planimetria dell’abitazione 5 e posizionamento dei PIR nelle stanze

Tuttavia, non sono stati installati 3 dispositivi PIR in tutte le abitazioni. L’abitazione numero 7, ad esempio, è priva del sensore collocato nella camera; l’abitazione numero 8 invece è sprovvista del PIR del bagno.

Per disporre di un dataset più vasto quindi per ottenere un monitoraggio dei soggetti migliore sono stati installati dei sensori all’interno degli interruttori della luce. Quest’ultimi danno in output 1 quando la luce viene accesa e 0 quando viene spenta.

Figura 13: interruttore della luce.

A differenza dei sensori PIR, che a causa delle planimetrie di alcune abitazioni non sono stati installati in maniera univoca, questi sensori sono presenti all'interno delle abitazioni nello stesso numero e posizionamento. Facendo riferimento all'abitazione numero 5 (home5), i sensori luci che sono stati presi in considerazione ai fini dell'analisi sono quelli evidenziati nella seguente planimetria, ovvero:

- Luce 1 installata nella stanza Cucina/Soggiorno;

- Luce 5 installata nel Bagno;

- Luce 6 installata nella Camera.

Figura 14: Planimetria casa 5 e posizionamento sensori luci.

4.2 Dati e segnali acquisiti

Da quanto descritto precedentemente è chiaro che l'architettura H@H è stata ideata e strutturata allo scopo di fornire servizi nelle abitazioni degli utenti, attraverso l'impiego di sensori "smart" che consentono di raccogliere e collezionare dati eterogenei secondo due livelli di immagazzinamento: Smart home e servizi Cloud.

Il primo livello consiste in un database locale organizzato in tabelle statiche, ossia una sorta di inventario dei singoli appartamenti che specifica come sono organizzati (dispositivi installati, utilizzatori) e tabelle dinamiche implementate per la raccolta dei dati "grezzi" provenienti dai singoli dispositivi.

Il secondo livello, d'altra parte, consiste in database Cloud che consente da un lato di acquisire periodicamente i dati per lo studio del progetto, dall'altro di renderli disponibili da remoto agli utilizzatori [16][17]. Grazie all'App "Salute" connessa con Cloud, realizzata per la gestione ed il controllo dei dispositivi domotici e biomedicali, sono stati acquisiti sia dati relativi ai parametri fisiologici che includono segnale elettrocardiografico (ECG), variabilità cardiaca (HR), temperatura timpanica, glicemia, pressione e livello di saturazione sanguigna e peso corporeo, che dati relativi a dispositivi Bticino. Quest'ultimi possono essere ricavati dalla cucina (ad esempio frigorifero), fornendo una strategia di dieta all'utente unificata ai servizi forniti dagli altri dispositivi connessi.

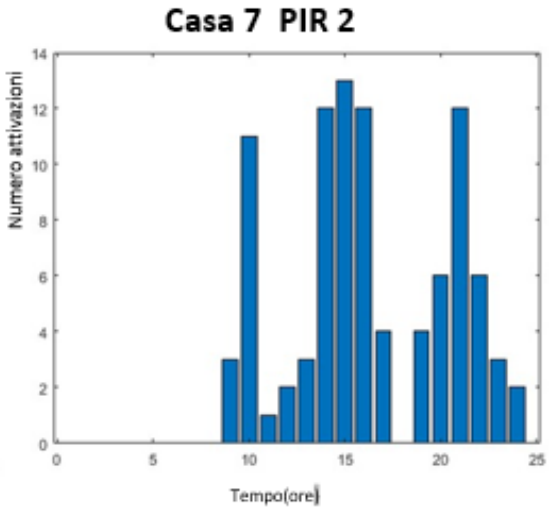

Figura 15: attivazioni registrate nell' arco di 24 ore del sensore PIR collocato nella camera da letto dell'abitazione 7.

In figura 15 sono mostrate le attivazioni provenienti dal sensore PIR della camera da letto durante una intera giornata. Nelle ascisse vengono indicate le ore della giornata, nelle ordinate viene riportato il numero di attivazioni in ore. Questi particolari dati sono molto importanti perché forniscono potenziali informazioni circa lo stile di vita degli individui. A partire da tali rilevamenti infatti si possono formulare espedienti al fine di migliorare la qualità di vita dei singoli coinvolti nel progetto e più in generale delle persone anziane [18][19].

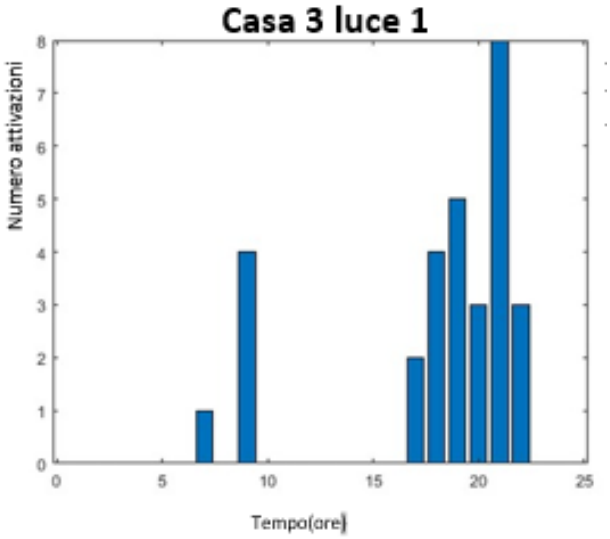

Figura 16: attivazioni registrate nell' arco di 24 ore del sensore luce collocato nel salotto dell'abitazione 3.

Il questionario è stato fondamentale ai fini di questa tesi. Il progetto H@H prevede che nell'arco di due mesi venga somministrato a tutti i soggetti un questionario. Quest'ultimo è composto da 10 domande, alcune riguardanti lo stato fisico e altre lo stato mentale. I soggetti hanno compilato il questionario a fine di ogni giornata del periodo che va dal 1° dicembre 2018 al 31 gennaio 2018. Di seguito viene riportato il modulo del questionario somministrato nella figura 16.

Casa 10				
				01 dicembre 2018
1. In generale il tuo stato di salute è:				
□ Ottimo	□ Molto buono	□ Buono	□ Nella media	□ Pessimo
2. Hai svolto attività moderate (e.g. lavare i piatti, pulire casa, cucinare)?				
□ Sì, con facilità	□ Sì, ma in modo limitato	□ No		
2.1 Il tuo stato di salute ti ha limitato nello svolgimento di queste attività?				
□ Sì, molto	□ Sì, un po'	□ No, affatto		
3. Hai svolto attività fisiche (camminare, salire più volte le scale, passeggiare)?				
□ Sì, molto	□ Sì, un po'	□ No, affatto		
4. Benessere mentale:				
□ Ottimo	□ Molto buono	□ Buono	□ Nella media	□ Pessimo
4.1 Situazioni di malessere emotivo hanno interferito con la tua vita abitudinaria?				
□ Sì, molto	□ Sì, un po'	□ No, affatto		
5. E' venuto a trovarti qualcuno?				
□ Sì	□ No			
5.1 Benessere fisico (stato generale di salute):				
□ Ottimo	□ Molto buono	□ Buono	□ Nella media	□ Pessimo
6. Situazioni di malessere fisico (dolori) hanno interferito con la tua vita abitudinaria?				
□ Sì, molto	□ Sì, un po'	□ No, affatto		

Figura 17: modulo del questionario somministrato.

4.3 Analisi dei dati

Considerando solamente i segnali domotici prelevati dalle abitazioni partecipanti al progetto, quali i dati provenienti dalle luci e dai sensori PIR, è stata data maggiore importanza agli ambienti "principali" delle case, ovvero la camera da letto, il bagno, ed infine il salotto.

Tuttavia, non tutti i questionari sono utili al nostro scopo. Ad esempio, il soggetto abitante "la casa 5" ha assegnato per tutto il periodo il voto 1: in questo modo è impossibile stabilire correlazioni tra accensioni e stato fisico/mentale quindi per questo motivo quest'ultima è stata scartata.

Sono stati analizzati quindi i dati provenienti dai sensori interessati di 4 abitazioni: casa 3, 4, 7 e 10. Le abitazioni numero 3 e 7 sono caratterizzate dalla presenza di due soggetti, al contrario casa 4 e 10 sono abitate da un solo inquilino. Di conseguenza per le prime due sono stati utilizzati due questionari appartenenti ai due soggetti.

L'analisi di tali segnali domotici è stata effettuata con la piattaforma di sviluppo Matlab, seguendo le fasi illustrate nel seguente diagramma di flusso:

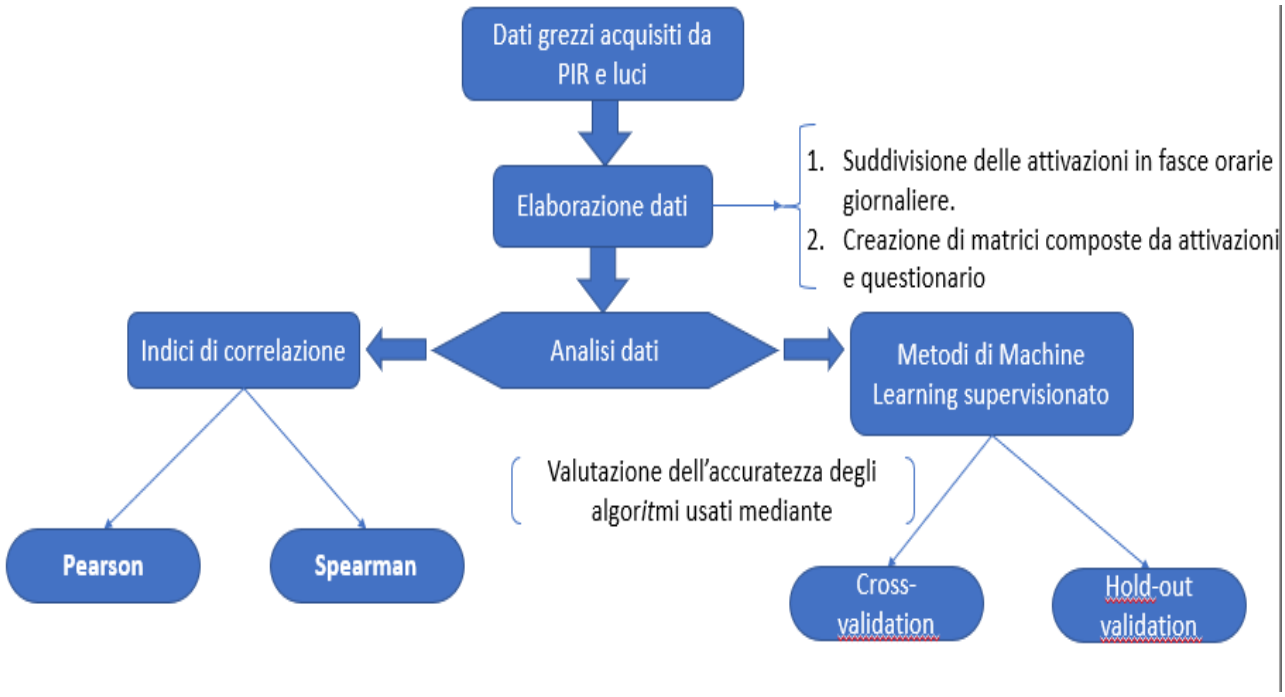

Figura 18: Diagramma di flusso dell'analisi dei dati.

Il programma sviluppato nell'ambiente Matlab prevede come primo step un controllo ricorsivo sui dati grezzi per verificare l'eventuale presenza di segnali spuri ed errori di acquisizione, ad esempio, l'identificazione di due uno o due zeri consecutivi è quindi una condizione non possibile. Dopo averli identificati, il programma elimina automaticamente questi dati. Questa operazione è stata eseguita solamente per i dati acquisiti dai sensori luci, a differenza di quest'ultimi i sensori PIR danno in output solo "1" quindi non è stato necessario filtrare i dati.

Il secondo step prevede la creazione delle matrici a cui è stato dato il nome "Alldays" ovvero delle matrici 24x390 dove le 24 righe rappresentano le ore del giorno e le colonne il numero di giorni di monitoraggio dei sensori PIR o dei sensori luci. Di conseguenza ogni casella rappresenta il numero di attivazioni che il sensore ha registrato in un'ora.

24x390 double

	1	2	3	4	5	6	7
1	0	0	0	1	0	0	3
2	0	0	0	1	0	0	3
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	9	8	4	6	6	1	3
10	3	7	7	4	4	4	10
11	8	7	4	0	4	7	5
12	7	4	2	0	7	4	1
13	5	7	6	6	3	3	7
14	5	2	4	5	3	0	3
15	0	2	1	1	1	0	1
16	2	2	2	6	5	4	3
17	6	4	8	0	0	4	3
18	0	5	5	0	0	0	2
19	5	10	5	8	10	3	5
20	5	2	1	1	3	9	4
21	6	6	4	0	0	4	7
22	3	7	5	5	1	4	8
23	3	4	3	0	0	4	0
24	1	1	0	3	0	1	1

Figura 19: Esempio di matrice “Alldays”.

Il terzo e ultimo step per l’analisi dei dati prevedeva la creazione delle “Matricicorrelazione” fondamentali per gli studi che saranno spiegati successivamente. Queste matrici sono state ottenute andando a selezionare all’interno delle matrici “Alldays” i giorni in cui è stato somministrato il questionario aggiungendo quest’ultimo così da ottenere matrici 62x25 dove le 62 righe rappresentano i giorni di monitoraggio dei PIR o sensori luci le colonne sono invece 25 in quanto alle 24 ore del giorno si è aggiunto il questionario.

Il questionario riportato in figura 19 è stato analizzato al fine di avere per ogni giorno un valore compreso tra 0 e 1 che indica lo stato di benessere mentale e fisico. Ai fini dell’analisi sono stati, dapprima, eliminati tutti i giorni in cui l’utente non ha risposto al questionario, ecco il motivo per cui non tutti i questionari presentavano 62 giorni di monitorazione. L’indice fisico e mentale è stato calcolato a partire dalle domande del questionario in particolare:

- Per il calcolo dell’indice fisico sono state considerate le domande 1, 2, 2.1, 3, 3.1 e la 5.1.
- Per il calcolo dell’indice mentale sono state considerate le domande 4 e 4.1 del questionario.

In entrambi i casi è stato assegnato un voto che va da 1 a 3 nel caso le risposte disponibili siano 3 e da 1 a 5 quando le risposte sono 5. È stata poi fatta la media di tali voti ed è stata in seguito normalizzata tra 0 e 1. Il risultato ottenuto è quindi un vettore in cui vengono riportati tutti i voti giorno dopo giorno.

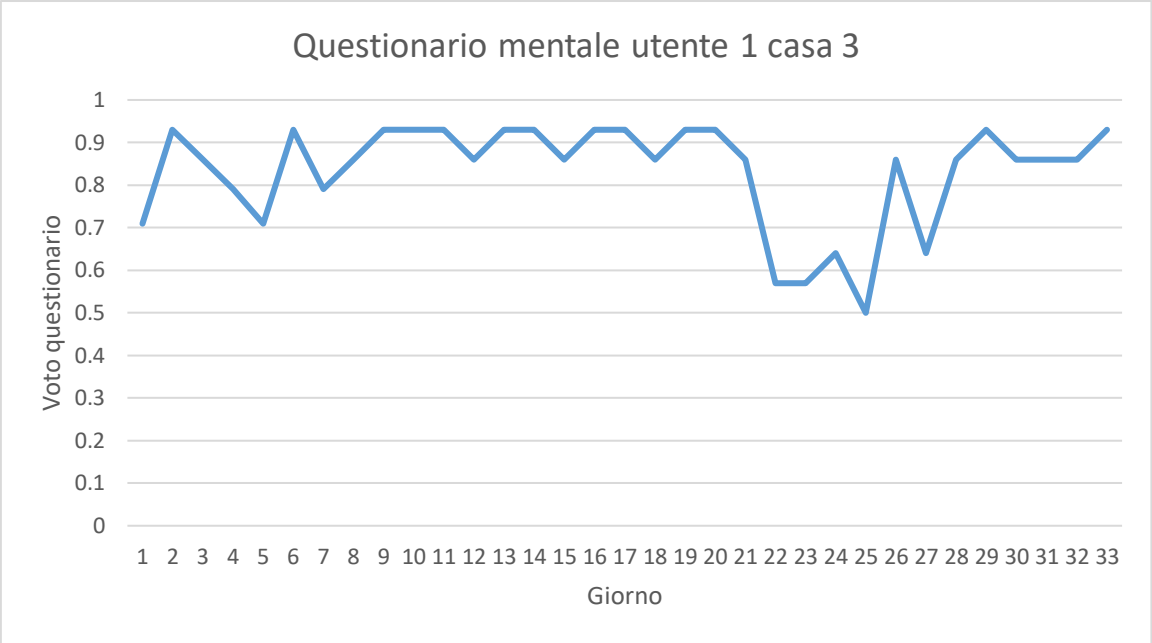

Figura 20: grafico dell'andamento dello stato mentale dell'utente 1 di casa 3.

Nella figura 20 viene riportato su un piano cartesiano l'andamento dello stato mentale dell'utente 1 della casa 3. Sull'asse delle ascisse vengono riportati i giorni in questo caso dal 1° dicembre al 2 gennaio 2018 mentre sull'asse delle ordinate il voto del questionario, come già detto da0 a 1.

4.3 Correlazioni attraverso Pearson e Spearman

I coefficienti di correlazione utilizzati sono stati due: Pearson e Spearman. Il Pearson è un indice di correlazione lineare e viene calcolato secondo la formula 1):

$$\boldsymbol{\rho_{A,B}} = \frac{\boldsymbol{cov(A,B)}}{\boldsymbol{\sigma_A \sigma_B}}.$$

L'indice di Pearson si ottiene quindi dal rapporto fra la covarianza di due variabili statistiche A e B, che è un indice di variabilità congiunta, e il prodotto della deviazione standard di A e di B. Come dicevamo, ρ è un indice di correlazione lineare e può assumere valori da -1 ad 1. Il valore zero indica che non c'è correlazione fra le due variabili, 1 indica la massima correlazione lineare in maniera proporzionale e -1 indica che la correlazione è inversamente proporzionale [20]. In altre parole, il Pearson ci dice quanto la dispersione dei punti delle due variabili è approssimabile con una retta, ma non ci dà informazioni sulla sensibilità ovvero sulla sua pendenza. Il coefficiente di correlazione Spearman, $\boldsymbol{\rho_s}$ o $\boldsymbol{r_s}$, è invece una versione non parametrica del Pearson e misura la forza e la direzione di una relazione monotona, meno restrittiva di una relazione lineare fra due variabili. Il calcolo di $\boldsymbol{\rho_s}$ avviene convertendo prima i valori in ranghi secondo la formula 2:

$$\rho_s = 1 - \frac{6\sum_i D_i^2}{N(N^2-1)}$$

Dove D_i rappresenta la differenza dei ranghi ovvero la differenza tra il rango della prima variabile e la il rango della seconda variabile; N è definito come l'insieme delle osservazioni. In conclusione, sono stati ritenuti significativi i valori che hanno riportato un indice di affidabilità statistica adeguato, nel nostro caso un p-value< 0,05 [20].

Le correlazioni sono state trovate mediante uno script Matlab, in particolare sono state cercate le correlazioni tra le attivazioni dei sensori PIR e sensori luci e i questionari somministrati andando a differenziare quest'ultimi nelle due parti ovvero stato fisico e mentale.

4.2 Metodi di Machine Learning

Per Machine Learning, o apprendimento automatico, si intendono quei metodi che consentono ad una macchina, intesa come computer, di apprendere dall'esperienza (i dati) ad eseguire degli specifici compiti senza che queste siano state esplicitamente e preventivamente programmate per farlo. Il Machine Learning iniziò a svilupparsi negli anni '60. Esso prende spunto da vari settori, come la statistica, il data mining e l'intelligenza artificiale, per portare allo sviluppo di quelli che sono gli algoritmi di Machine Learning e che permettono alla macchina di allenare un classificatore a fare previsioni su dati sconosciuti sulla base di una serie di dati conosciuti. Questi studi di apprendimento automatico oggi rappresentano degli strumenti molto utilizzati in ambito industriale e della ricerca per analisi di Big Data con applicazione nei settori più disparati [22] [23].

Tipicamente possiamo dividere le modalità di apprendimento della macchina in tre categorie: apprendimento supervisionato, non supervisionato ed apprendimento con rinforzo. L'apprendimento supervisionato si basa sul concetto di addestramento, cioè diamo all'algoritmo un insieme di coppie di dati input-output conosciute. Per supervisionato si intende quindi che, nell'insieme di dati di addestramento, ad ogni dato input corrisponde un output, fornito da un supervisore (una persona), che è il risultato corretto di quell'input. Da questi dati la macchina elabora un modello predittivo per i prossimi dati input di cui non si conosce l'output [22] [23] [24]. In questa maniera, gli algoritmi di Machine Learning potranno analizzare i dati in maniera più veloce e risolvere problemi o compiti in modo autonomo basandosi sulle esperienze precedenti. Un algoritmo di apprendimento supervisionato può essere utilizzato in vari settori come in quello medico (può fare una previsione dello scatenarsi di particolari crisi come pandemie e insorgenza di tumori in base a dati biometrici del passato), identificazione vocale (migliora il sistema in base agli ascolti audio passati) o l'identificazione della scrittura manuale (migliorata grazie all'analisi di centinaia di documenti scritti).

L'apprendimento non supervisionato, invece, si distingue dal primo perché nei dati di allenamento del classificatore non si forniscono agli input degli output noti, probabilmente perché non sappiamo esattamente quale sia la risposta corretta per quei dati. Abbiamo quindi una serie di input non classificati né decodificati a priori. Così, in questo caso, è la macchina stessa a cercare fra i dati delle caratteristiche comuni e ad organizzarli in classi in modo da avere un modello per fare previsioni sui

dati futuri [23] [24] [25]. L'apprendimento non supervisionato è impiegato per tecniche di comprensione dei dati. Tecniche di questo genere sono utilizzate dai motori di ricerca che, partendo da una o più parole chiave, creano liste di risultati (le SERP Search Engine Results Page) ritenute attinenti alla ricerca effettuata. L'ultimo metodo di apprendimento è quello con rinforzo. In questo caso la macchina migliora le proprie performance grazie a delle funzioni di rinforzo che, interagendo con l'ambiente circostante, verificano se effettivamente la risposta prevista ad un determinato input si è rivelata corretta. Il rinforzo è fornito da dei sensori presenti nell'ambiente che valutano con dei feedback l'output previsto. Quindi, inizialmente la macchina fa previsioni in modo casuale per poi orientarsi sui modelli predittivi che hanno avuto un maggior numero di feedback positivi. Ad esempio, potremmo trovare questo tipo di apprendimento nell'addestramento di un sistema per il gioco degli scacchi [23] [24] [25].

In questa tesi il Machine Learning è stato impiegato per trovare il modello più accurato nel fare previsioni sullo stato di benessere fisico e mentale di un soggetto partendo dalle attivazioni dei sensori PIR e sensori luci. Lo strumento adoperato è il Classification Learner.

Il Classification Learner è un toolbox di MATLAB che utilizza metodi di apprendimento supervisionato per allenare il proprio modello predittivo. Nel nostro specifico caso i dati forniti al classificatore sono stati i valori delle i dati acquisiti dai sensori PIR e sensori luci e il questionario sullo stato fisico e mentale. I primi rappresentano i predittori, e sono, dunque, i dati forniti come input al classificatore. L'apprendimento è supervisionato in quanto ai predittori corrispondono delle risposte, dati output, che nel nostro caso sono i valori del questionario fisico/mentale. Questo rappresenta una prova empirica sullo stato fisico e mentale dei soggetti durante il periodo di misurazione.

Si prosegue con la scelta del modello. Inizialmente si sono svolti dei test direttamente sulla piattaforma del Classification Learner che hanno portato alla scelta dei modelli predittivi che usavano i seguenti algoritmi di apprendimento supervisionato:

1. Decision Trees;

2. SVM (Support Vector Machines);

3. K-nearest neighbors (KNN).

Ogni algoritmo estrae dai dati delle caratteristiche, e quindi allena il modello, in maniera differente. Sulla base di questo il modello risulterà più o meno efficace nel fare previsioni su dati futuri. L'efficacia del modello sarà poi quantificata calcolando la sua accuratezza, accuracy, che vedremo in seguito. Vediamo, quindi, il significato di questi quattro algoritmi, tutti di apprendimento supervisionato.

1. Il Decision Tree seleziona, in base ai dati, delle caratteristiche per creare un nodo decisionale che divida il dataset in più gruppi o classi. Così, al primo passo il dataset di partenza è diviso in due rami da quello che viene definito il Root Node o Radice. L'operazione di ramificazione continua portando a classificare i dati in modo sempre più specifico ad ogni passo. Questo algoritmo si mostra utile e veloce nell'analisi del dataset. È inoltre facile da interpretare anche con rappresentazioni grafiche [26].

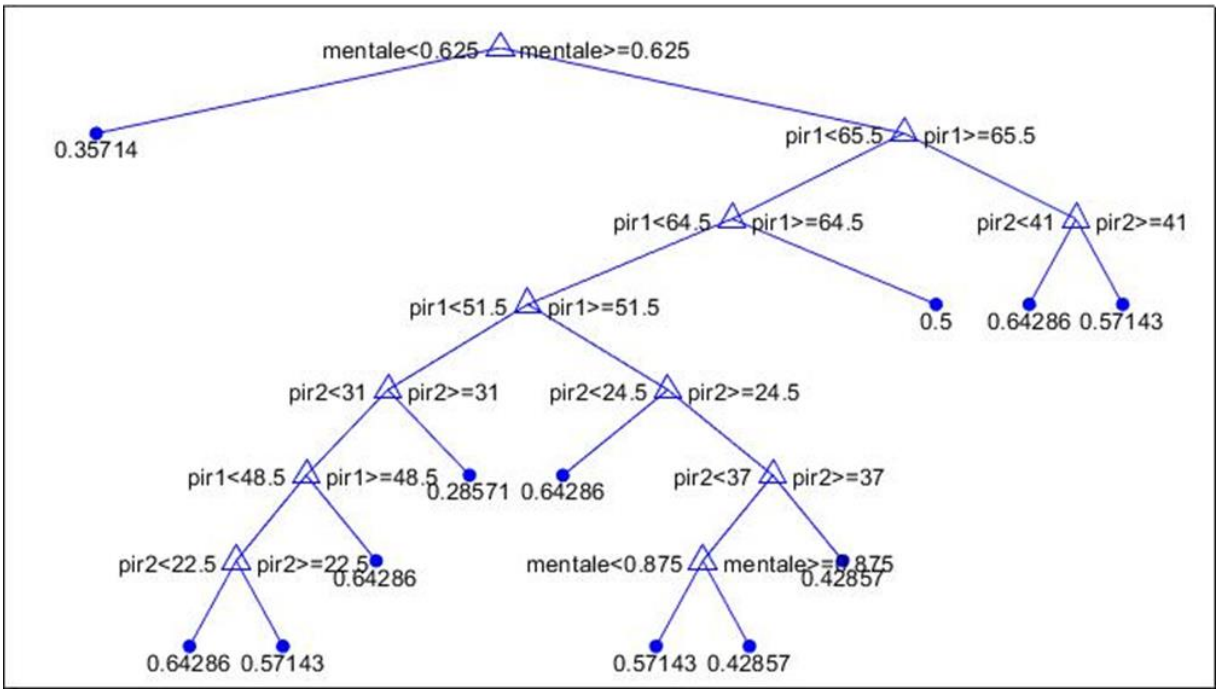

Figura 21: esempio di albero decisionale ottenuto dalla relazione tra le attivazioni del PIR 1 di casa 5 e il questionario mentale.

2. Il Support Vector Machine si basa sull'idea di trovare un 'iperpiano' che divida al meglio un set di dati in due classi. L'iperpiano (sottospazio lineare di dimensione (n-1) rispetto allo spazio di dimensione n in cui è contenuto) è definito da un vettore di supporto che dipende dai dati che si stanno analizzando. L'operazione viene poi iterata. Il metodo è difficile da interpretare ma si presta bene per problemi di riconoscimento binari (vero o falso, 0 o 1, riconoscimento vocale o di immagini e la computer vision) [27].

3. Il funzionamento del KNN si basa sulla somiglianza fra dati. La somiglianza è calcolata spesso in riferimento alla distanza euclidea fra punti dati e in base a questa seleziona delle classi. Così l'algoritmo calcola la distanza di un nuovo data point da tutti gli altri data points di allenamento. Quindi seleziona i punti K, dove K può essere un qualsiasi numero intero, e infine assegna i data points alla classe che riscontra la maggior somiglianza [28].

I metodi sopra elencati si dividono a loro volta in sottocategorie, di queste sono state scelte quelle che meglio si prestano ovvero:

- FINE KNN
- FINETREE
- MEDIUMTREE
- COARSETREE
- BUGGEDTREE
- LINEARSVM

Per ogni modello allenato è stata estratta la funzione dal toolbox per adoperarla su un apposito script in MATLAB che iterasse l'operazione per tutti i soggetti in modo automatico e che consentisse di generare previsioni su nuovi dati.

Come era stato prima accennato, l'efficacia del modello predittivo è quantificabile calcolando la sua accuratezza. L'accuratezza viene stabilita secondo la formula 3):

$$Accuracy = \frac{TP + TN}{TN + FP + FN + TP}$$

Essa è data quindi dal rapporto tra le predizioni esatte (true positive e true negative) e le predizioni totali effettuate (true negative, false positive, false negative e true positive). [29]

Sarà quindi un numero compreso fra 0 ed 1 che, riportato in percentuale, esprime quanto è affidabile la previsione fatta dalla macchina. Ovviamente tanto più l'accuratezza calcolata si avvicina ad 1 (in percentuale 100%) migliore sarà la previsione dell'algoritmo.

Ciò che deve assolutamente essere evitato, utilizzando queste tecniche è il fenomeno dell'Overfitting. Nel machine learning l'Overfitting è il rischio di sovradattamento durante il processo di apprendimento: ovvero consiste nell'individuazione da parte dell'algoritmo di false regolarità. Questo fenomeno fa sì che alcuni dati mostrano una relazione tra un attributo irrilevante X e il risultato finale Y. Apparentemente il modello sembra perfetto tuttavia, si tratta di rumori che sporcano i dati e fuorviano l'apprendimento induttivo della macchina. Detto in modo più semplice, l'algoritmo si adatta (fitting) troppo bene (over) ai dati di training perdendo in generalità. L'Overfitting può avere diverse cause:

- La macchina o il supervisore prende in considerazione nell'insieme di addestramento (training set) degli attributi X irrilevanti con la decisione finale Y.
- I dati di training sono insufficienti.
- I dati di training non sono rappresentativi della realtà [30].

4.3 Metodo di valutazione: **Cross-Validation**

Uno dei due metodi utilizzati per valutare l'accuratezza degli algoritmi utilizzati è il metodo della Cross-Validation. Questa è una tecnica statistica utilizzabile in presenza di una buona numerosità del campione osservato. In particolare, la convalida incrociata cosiddetta k-fold consiste nella suddivisione dell'insieme di dati totale in k parti di uguale numerosità e, a ogni passo, la kª parte dell'insieme di dati viene a essere quella di convalida, mentre la restante parte costituisce sempre l'insieme di addestramento. [31]

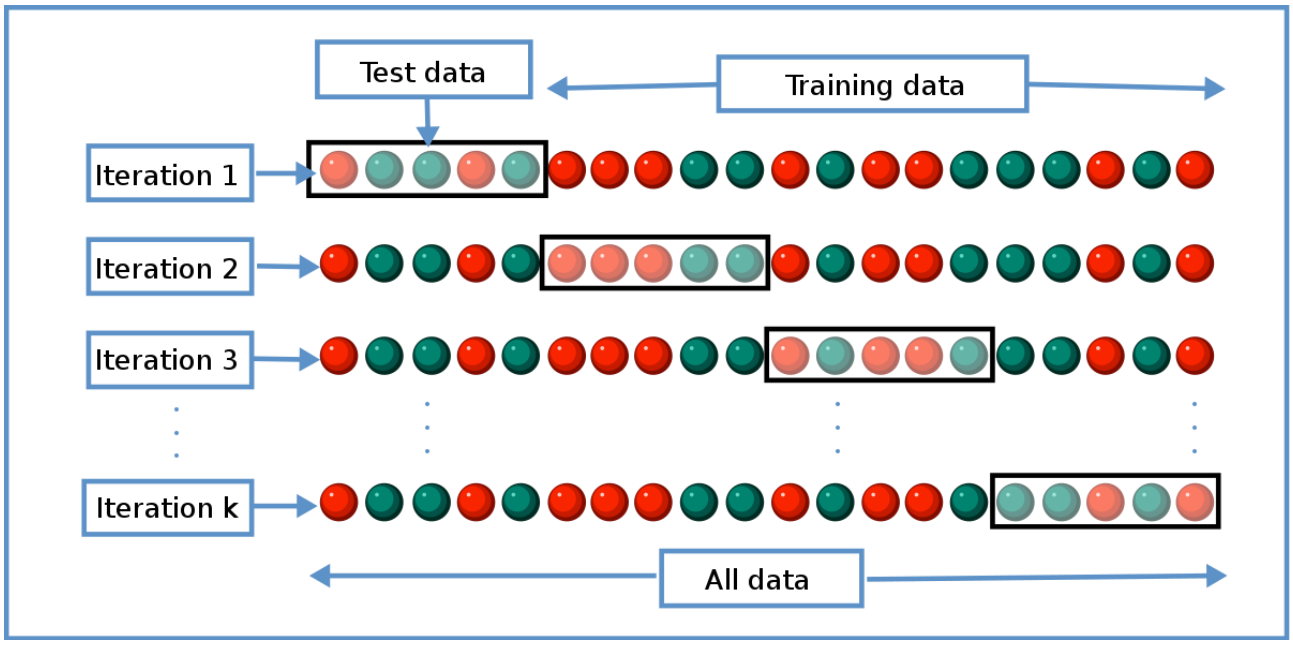

Figura 22: schema di funzionamento del metodo Cross-Validation

Dunque, in questo progetto l'accuratezza è stata calcolata con un metodo di cross validation a 10 fold. Questo consiste nella suddivisione del dataset totale in k parti di uguale dimensioni, 10 nel nostro caso, e, ad ogni passo, la k-esima parte del dataset è usata come dataset per la validazione, mentre le restanti parti costituiscono il dataset per il training. Il procedimento si ripete finché tutte le parti sono usate sia per l'allenamento che per la validazione.

Dapprima sono stati creati dei set di dati per ogni abitazione, dopo aver avviato il tool Classification Learner di Matlab sono state caricate su di esso le "matrici correlazione". Data la loro struttura è stato possibile dare in input al classification learner le prime 24 colonne (attivazioni PIR o luci) ed è stata esclusa dalla selezione l'ultima colonna ovvero quella del questionario da predirre. Sono stati in seguito selezionati gli algoritmi sopra elencati e è stato addestrato il programma mediante k valori presi dal questionario in seguito quest'ultimo predice gli (n-k) valori rimanenti.

4.4 Metodo di valutazione: **Hold-out Validation**

Nel metodo hold-out, vengono creati casualmente due dataset, d_0 e d_1, generalmente chiamati dataset di allenamento e dataset di test. La dimensione di ciascuno dei dataset è arbitraria, sebbene in genere il set di test sia più piccolo del set di allenamento. Quindi alleniamo (costruiamo un modello) su d_0 e testiamo (valutiamo le sue prestazioni) su d_1 [32].

Split 1	Sample 1	Sample 2	Sample 3	...	Sample n
Split 2	Sample 1	Sample 2	Sample 3	...	Sample n
Split 3	Sample 1	Sample 2	Sample 3	...	Sample n
...	Sample 1	Sample 2	Sample 3	...	Sample n
Split n	Sample 1	Sample 2	Sample 3	...	Sample n
	TRAINING DATA			TEST DATA	

Figura 23: Schema di funzionamento dell' Hold-out Validation

L'iter eseguito per questo metodo nella tesi è il seguente: inizialmente sono state estratte le funzioni dal Classification Learner di ciascuno degli algoritmi utilizzati, in seguito è stato creato uno script Matlab che ci ha permesso di eseguire questo tipo di valutazione.

Mediante un ciclo for è stato possibile escludere in ogni ciclo un valore dal questionario e allenare l'algoritmo o funzione, Fitcknn in questo caso, sui restanti andando poi a predirre il valore escluso.

In seguito, per valutare l'accuratezza è stata introdotta la funzione RMSE, in esempio chiamata ''My_RMSE'' definita come nella seguente formula 5):

$$\mathrm{RMSE}_{fo} = [\sum_{i=1}^{N} (z_{f_i} - z_{o_i})^2 / N]^{1/2}$$

Definita come **Root Mean Square Error** è la differenza tra il valore predetto dal questionario (z_{fi}=r) e il valore reale del questionario (z_{oi}=predit) dove N indica il numero totale delle osservazioni. Da ciò si evince che minore è l'RMSE calcolato tanto migliore sarà la predizione dell'algoritmo in quanto per definizione l'RMSE è la differenza tra il valore predetto dell'algoritmo e il valore vero del questionario.

5. Risultati

In questa sezione saranno mostrati ed esaminati i risultati ottenuti tramite indici di correlazione, tra il questionario e il numero di attivazioni, mediante tecniche di Machine learning che sono state validate prima tramite cross-validation e poi tramite hold-out validation.

5.1 Risultati delle correlazioni

In questa sezione mostriamo le correlazioni mediante gli indici Pearson, indicato come RHO_P, e Spearman, indicato con RHO_S, tra le attivazioni dei sensori PIR e i questionari somministrati ai soggetti. In particolare, saranno riportati i risultati ottenuti dai dati raccolti dai sensori PIR 1 e 2 in quanto presenti in tutte le abitazioni.

Gli indici sono stati calcolati associando separatamente a entrambi i PIR una volta il questionario fisico e una volta il questionario mentale.

Nelle tabelle sottostanti sono riportati per tutte le abitazioni i relativi indici di correlazione e per ognuno di questi è specificato il valore "rho" e il p-value anche quest'ultimo indica con PVAL_P se riferito all'indice Pearson e con PVAL_S a indice Spearman.

INDICI CALCOLATI SU ATTIVAZIONI PIR 1				
	FISICO			
Casa	RHO_P	PVAL_P	RHO_S	PVAL_S
4	-0,01	0,97	0,01	0,96
5	-0,03	0,81	-0,08	0,53
7	-0,10	0,43	-0,15	0,23
10	0,06	0,63	0,11	0,41
3	0,23	0,21	0,08	0,68
Media	0,34	_	0,38	_

Tabella 1: indici di correlazione tra le attivazioni del PIR 1 e il questionario fisico.

INDICI CALCOLATI SU ATTIVAZIONI PIR 1				
	MENTALE			
Casa	RHO_P	PVAL_P	RHO_S	PVAL_S
4	0,03	0,94	-0,01	0,94
5	_	_	_	_
7	0,00	1,00	-0,13	0,31
10	-0,09	0,49	-0,08	0,51
3	0,57	_	0,42	0,02
Media	0,48	_	0,53	_

Tabella 2: indici di correlazione tra le attivazioni del PIR 1 e il questionario mentale.

INDICI CALCOLATI SU ATTIVAZIONI PIR 2				
	FISICO			
Casa	RHO_P	PVAL_P	RHO_S	PVAL_S
4	-0,08	0,57	0,02	0,88
5	-0,16	0,21	-0,17	0,19
7	0,05	0,71	-0,01	0,95
10	-0,04	0,75	-0,01	0,95
3	0,13	0,47	0,17	0,37
Media	-0,28	_	-0,30	_

Tabella 3: indici di correlazione tra le attivazioni del PIR 2 e il questionario fisico.

INDICI CALCOLATI SU ATTIVAZIONI PIR 2				
	MENTALE			
Casa	RHO_P	PVAL_P	RHO_S	PVAL_S
4	-0,24	0,09	-0,30	0,04
5	_	_	_	_
7	0,03	0,84	-0,08	0,56
10	-0,03	0,84	-0,03	0,81
3	0,38	0,04	0,37	0,04
Media	-0,15	0,01	0,53	_

Tabella 4: indici di correlazione tra le attivazioni del PIR 2 e il questionario mentale.

Esaminando le tabelle 1, 2, 3 e 4 notiamo che:

- Le correlazioni tra le attivazioni PIR e i questionari riguardanti lo stato fisico sono molto basse, non troviamo infatti valori di rho che si avvicinano ad 1 ovvero al valore di perfetta correlazione, inoltre tutti i p-value sono minori di 0,05 di conseguenza non vi è alcuna significatività statistica.

- Le correlazioni tra le attivazioni PIR e i questionari riguardanti lo stato mentale sono molto basse. In particolare, gli indici di correlazione più elevati trovati riguardano l'abitazione numero 3 e il PIR 2 dove si registra un indice Pearson di 0,38 e un indice Spearman di 0,37 a cui sono associati p-value pari a 0,04.

Si può concludere che attraverso le correlazioni con i coefficienti Pearson e Spearman è difficile valutare con efficacia la dipendenza fra le attivazioni dei sensori PIR e i questionari al fine di fare previsioni accurate sullo stato di comfort fisico e mentale del soggetto partendo dai dati raccolti dai sensori PIR. Per questo motivo si è dunque deciso di cambiare approccio metodologico e di sfruttare metodi di Machine Learning.

5.2 Risultati del ML

Come era stato prima trattato, è stata svolta un'analisi dei dati mediante l'utilizzo del Machine Learning quindi tramite dei modelli predittivi. L'efficacia di questi modelli è stata quantificata calcolando l'accuratezza di quest'ultimi.

I metodi di validazione utilizzati in questo studio, come già riportato precedentemente, sono due ovvero la cross-validation e l'hold-out validation. In questa sezione saranno esaminati i risultati ottenuti mediante i metodi sopra citati e nel medesimo ordine. Le accuratezze sono state calcolate associando separatamente ai sensori PIR e ai sensori luci una volta il questionario fisico e una volta il questionario mentale. Saranno di seguito riportate in tabella le accuratezze relative prima ai sensori PIR e poi ai sensori luci.

ACCURATEZZE								
casa 3								
	MENTALE				FISICO			
	pir1		pir2		pir1		pir2	
	Utente 1	Utente 2	Utente 1	Utente 2	Utente 1	Utente 2	Utente 1	Utente 2
FINE KNN	0,47	0,87	0,47	0,87	0,73	1,00	0,91	1,00
FINETREE	0,39	0,90	0,44	0,86	0,55	1,00	0,55	1,00
MEDIUMTR	0,46	0,90	0,31	0,86	0,55	1,00	0,55	1,00
COARSETR	0,41	0,90	0,49	0,86	0,55	1,00	0,55	1,00
BAGGEDTR	0,47	0,80	0,41	0,89	0,55	1,00	0,64	1,00
LINEAR SV	0,49	0,00	0,56	0,00	0,00	0,00	0,00	0,00
media	0,45	0,87	0,45	0,87	0,58	0,83	0,64	0,83

Tabella 5: accuratezze calcolate mediante cross-validation in casa 3 relative a sensori PIR.

ACCURATEZZE				
Casa 4				
	MENTALE		FISICO	
	PIR1	PIR2	PIR1	PIR2
FINE KNN	0,73	0,62	0,18	0,16
FINETREE	0,80	0,76	0,18	0,14
MEDIUMT	0,64	0,76	0,18	0,14
COARSETF	0,73	0,78	0,20	0,14
BAGGEDTI	0,78	0,76	0,22	0,16
LINEAR SV	0,36	0,47	0,00	0,00

Tabella 6: accuratezze calcolate mediante cross-validation in casa 4 relative a sensori PIR.

ACCURATEZZE								
Casa 7								
	Mentale				Fisico			
	PIR 1		PIR 2		PIR 1		PIR 2	
	Utente 1	Utente 2	Utente 1	Utente 2	Utente 1	Utente 2	Utente 1	Utente 2
FINE KNN	0,79	0,75	0,85	0,82	0,15	1	0,26	1
FINETREE	0,67	0,74	0,77	0,80	0,26	1	0,21	1
MEDIUMTREE	0,65	0,64	0,65	0,70	0,26	1	0,21	1
COARSETREE	0,73	0,64	0,69	0,79	0,29	1	0,26	1
BAGGEDTREE	0,83	0,77	0,81	0,90	0,23	1	0,27	1
LINEAR SVM	0,00	0,00	0,00	0,00	0,24	0	0,31	0

Tabella 7: accuratezze calcolate mediante cross-validation in casa 7 relative a sensori PIR.

ACCURATEZZE						
Casa 10						
	MENTALE			FISICO		
	PIR1	PIR2	PIR3	PIR1	PIR2	PIR3
FINE KNN	0,88	0,88	0,75	1,00	1,00	1,00
FINETREE	0,79	0,63	0,75	1,00	1,00	1,00
MEDIUMTREE	0,83	0,75	0,83	1,00	1,00	1,00
COARSETREE	0,88	0,67	0,88	1,00	1,00	1,00
BAGGEDTREE	0,75	0,79	0,79	1,00	1,00	1,00
LINEAR SVM	0,00	0,00	0,00	0,00	0,00	0,00

Tabella 8: accuratezze calcolate mediante cross-validation in casa 10 relative a sensori PIR.

In riferimento alla tabella 5 relativa all'abitazione 3 gli algoritmi che hanno ottenuto accuratezze maggiori dell'80% sono il FineKNN e gli alberi decisionali raggiungendo accuratezze molto buone circa del 90% questo trend cambia completamente per le accuratezze relative al questionario mentale dove è molto frequente il fenomeno dell'Overfitting portando valori di accuratezza pari ad 1 e che quindi sono state scartate.

In generale, per quel che concerne i sensori PIR, possiamo notare una netta differenza tra le accuratezze calcolate a partire dal questionario mentale e quelle del questionario fisico, è evidente che le prime sono maggiori delle seconde. I modelli predittivi hanno generato buone accuratezze per il questionario sullo stato mentale soprattutto grazie agli algoritmi degli alberi decisionali ma molto basse per il questionario sullo stato fisico con tanti fenomeni di Overfitting.

ACCURATEZZE												
casa 3												
	MENTALE						FISICO					
	luce1		luce5		luce6		luce1		luce5		luce6	
	Utente 1	Utente 2	Utente 1	Utente 2	Utente 1	Utente 2	Utente 1	Utente 2	Utente 1	Utente 2	Utente 1	Utente 2
FINE KNN	0,46	0,84	0,39	0,79	0,36	0,76	0,82	1,00	0,55	1,00	0,91	1,00
FINETREE	0,51	0,89	0,39	0,84	0,37	0,90	0,55	1,00	0,55	1,00	0,55	1,00
MEDIUMTF	0,47	0,89	0,47	0,84	0,44	0,90	0,55	1,00	0,55	1,00	0,55	1,00
COARSETR	0,59	0,89	0,53	0,87	0,42	0,90	0,55	1,00	0,55	1,00	0,55	1,00
BAGGEDTR	0,53	0,89	0,53	0,87	0,47	0,87	0,64	1,00	0,27	1,00	0,55	1,00
LINEAR SV	0,56	0,90	0,58	0,90	0,59	0,90	0,00	0,00	0,00	0,00	0,00	0,00
MEDIA	0,52	0,88	0,48	0,85	0,44	0,87	0,62	1,00	0,49	1,00	0,62	1,00

t

Tabella 9: accuratezze calcolate mediante cross-validation in casa 3 relative a sensori luci.

ACCURATEZZE						
Casa 4						
	MENTALE			FISICO		
	luce1	luce5	luce6	luce1	luce5	luce6
FINE KNN	0,64	0,71	0,69	0,26	0,20	0,16
FINETREE	0,80	0,78	0,62	0,14	0,20	0,18
MEDIUMTREE	0,80	0,67	0,78	0,14	0,20	0,18
COARSETREE	0,73	0,82	0,76	0,20	0,22	0,18
BAGGEDTREE	0,78	0,80	0,78	0,16	0,20	0,14
LINEAR SVM	0,36	0,40	0,42	0,29	0,22	0,25

tabella 10: accuratezze calcolate mediante cross-validation in casa 4 relative a sensori luci.

ACCURATEZZE												
Casa 7												
	MENTALE						FISICO					
	luce1		luce5		luce6		luce1		luce5		luce6	
	Utente 1	Utente 2	Utente 1	Utente 2	Utente 1	Utente 2	Utente 1	Utente 2	Utente 1	Utente 2	Utente 1	Utente 2
FINE KNN	0,73	0,84	0,75	0,79	0,73	0,79	0,36	1	0,16	1	0,18	1
FINETREE	0,83	0,84	0,63	0,69	0,67	0,75	0,26	1	0,26	1	0,21	1
MEDIUMTREE	0,77	0,79	0,71	0,72	0,77	0,77	0,26	1	0,26	1	0,21	1
COARSETREE	0,79	0,87	0,65	0,72	0,65	0,80	0,24	1	0,21	1	0,21	1
BAGGEDTREE	0,79	0,84	0,71	0,82	0,77	0,85	0,29	1	0,27	1	0,24	1
LINEAR SVM	0,00	0,00	0,00	0,00	0,00	0,00	0,29	0	0,23	0	0,26	0

Tabella 11: accuratezze calcolate mediante cross-validation in casa 4 relative a sensori luci.

ACCURATEZZE						
Casa 10						
	MENTALE			FISICO		
	luce1	luce5	luce6	luce1	luce5	luce6
FINE KNN	0,79	0,88	0,79	1,00	1,00	1,00
FINETREE	0,83	0,83	0,88	1,00	1,00	1,00
MEDIUMTREE	0,83	0,83	0,88	1,00	1,00	1,00
COARSETREE	0,83	0,79	0,75	1,00	1,00	1,00
BAGGEDTREE	0,88	0,88	0,83	1,00	1,00	1,00
LINEAR SVM	0,00	0,00	0,00	0,00	0,00	0,00

Tabella 12: accuratezze calcolate mediante cross-validation in casa 4 relative a sensori luci.

In riferimento alla tabella 11 relativa all'abitazione n. 7 gli algoritmi che hanno ottenuto accuratezze maggiori dell'80% sono il FineKNN, il Coarse Tree e il Bugged Tree raggiungendo accuratezze molto buone del 87% questo trend cambia completamente per le accuratezze relative al questionario mentale dove è molto frequente il fenomeno dell'Overfitting portando valori di accuratezza pari ad 1 e che quindi sono state scartate.

In generale, come per i sensori PIR, anche per i sensori luci le accuratezze migliori riguardano il questionario mentale. In questo caso il modello che ha restituito le predizioni più accurate è stato il Bugged Tree con massimi del 90%. Il fenomeno dell'Overfitting risulta essere molto comune nell'analisi del questionario fisico anche a partire dalle attivazioni dei sensori luce. Questo problema è comune soprattutto alle luci 1 e 6.

È possibile concludere che le accuratezze calcolate mediante il metodo della cross-validation, seppur buone nella predizione del questionario mentale sono scarse sulla predizione del questionario fisico, per questo motivo è stata eseguita una validazione mediante hold-out.

Come già discusso nel capitolo 4.7 è stata eseguita una validazione degli algoritmi mediante hold-out in seguito per ogni validazione sono stati calcolati gli RMSE tra il valore reale del questionario e quello predetto. Di seguito verranno riportati due esempi degli RMSE calcolati sotto forma di tabella (tabella 13 e 14), in particolare i valori relativi al PIR 1 associato al questionario fisico e alla luce 1 associati al questionario mentale.

PIR 1 FISICO						
	RMSE					
	FINE KNN	FINE TREE	MEDIUM TREE	COARSE TREE	BAGGED TREE	LINEAR SVM
Casa 3 U1	**0,13**	0,00	0,11	0,11	0,11	0,11
Casa 3 U2	0,06	0,06	0,04	0,04	0,04	0,04
Casa 4	0,07	0,10	0,10	0,07	0,07	0,07
Casa 7 U1	0,16	0,18	0,12	0,12	0,12	0,12
Casa 7 U2	0,10	0,13	0,09	0,09	0,09	0,09
Casa 10	0,22	0,21	0,07	0,07	0,07	0,07

Tabella 13: RMSE ottenuto tra la predizione del Machine Learning e il valore vero del questionario fisico.

LUCE 1 MENTALE						
	RMSE					
	FINE KNN	FINE TREE	MEDIUM TREE	COARSE TREE	BAGGED TREE	LINEAR SVM
Casa 3 U1	0,12	0,07	0,08	0,15	0,10	0,20
Casa 3 U2	0,15	0,07	0,09	0,18	0,10	0,21
Casa 4	0,11	0,06	0,17	0,11	0,08	0,15
Casa 7 U1	0,11	0,06	0,06	0,11	0,08	0,15
Casa 7 U2	0,11	0,06	0,06	0,11	0,08	0,15
Casa 10	0,11	0,06	0,06	0,11	0,08	0,15

Tabella 14: RMSE ottenuto tra la predizione del Machine Learning e il valore vero del questionario mentale.

Per poter visualizzare in modo ottimale i dati raccolti sono stati creati degli istogrammi: nelle ascisse vengono riportati gli algoritmi utilizzati e nelle ordinate è indicato il valore dell'RMSE calcolato inoltre ogni barra indica l'utente o la casa a cui ci si riferisce. Di seguito verranno prima riportati gli istogrammi relativi alle accensioni dei sensori PIR e infine quelli relativi alle accensioni dei sensori luce.

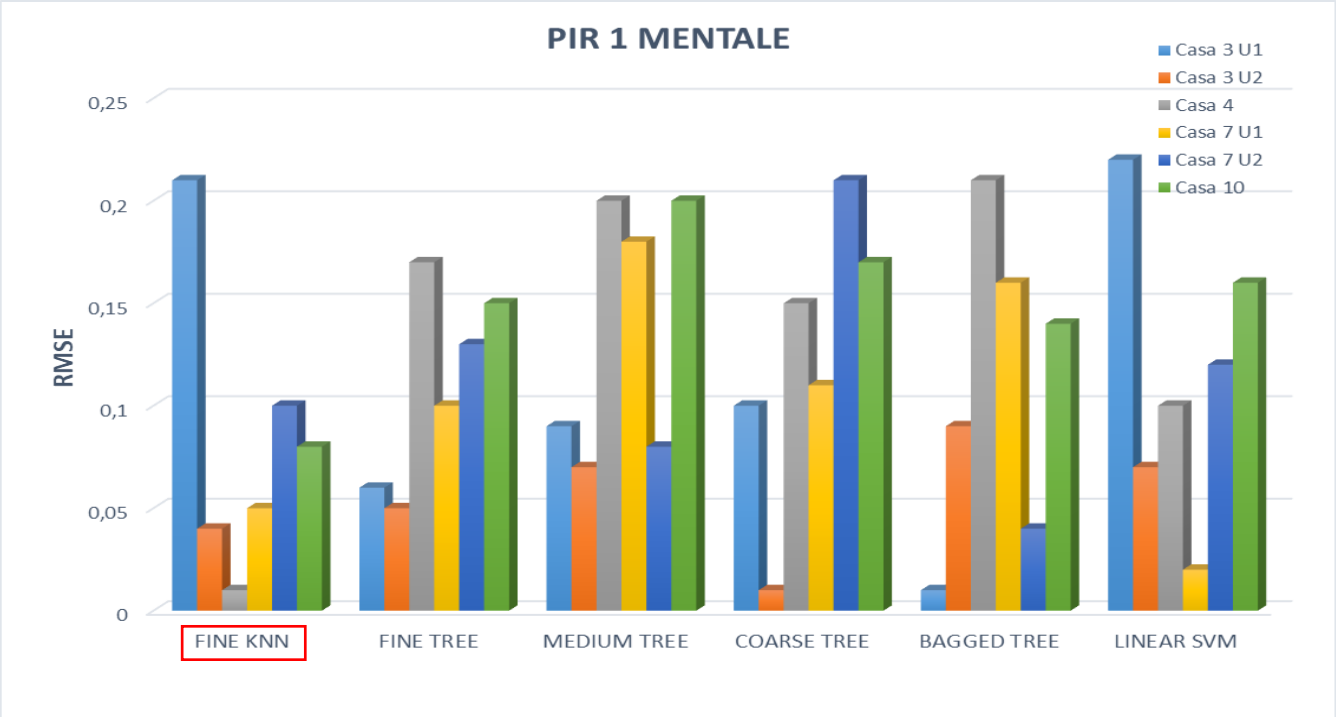

Figura 24: istogramma degli RMSE che esprimono l'accuratezza dei metodi di Machine Learning per predirre il questionario mentale a partire dai dati del PIR 1.

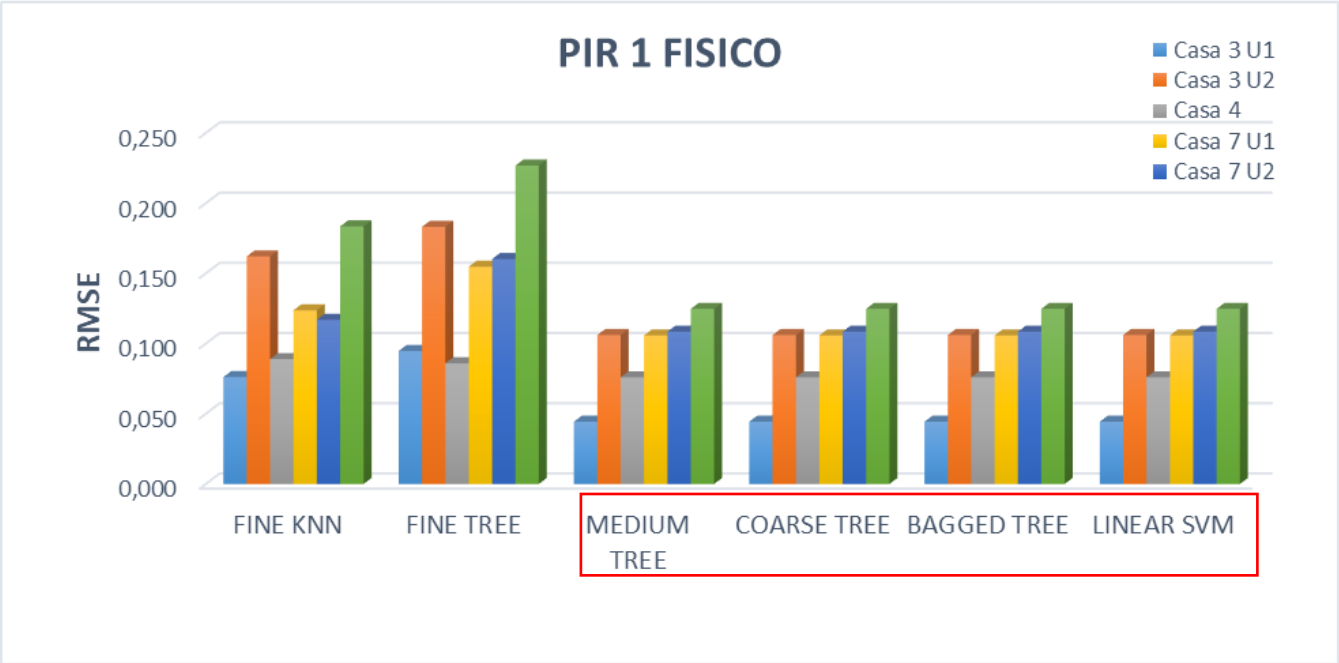

Figura 25: istogramma degli RMSE che esprimono l'accuratezza dei metodi di Machine Learning per predirre il questionario fisico a partire dai dati del PIR 1.

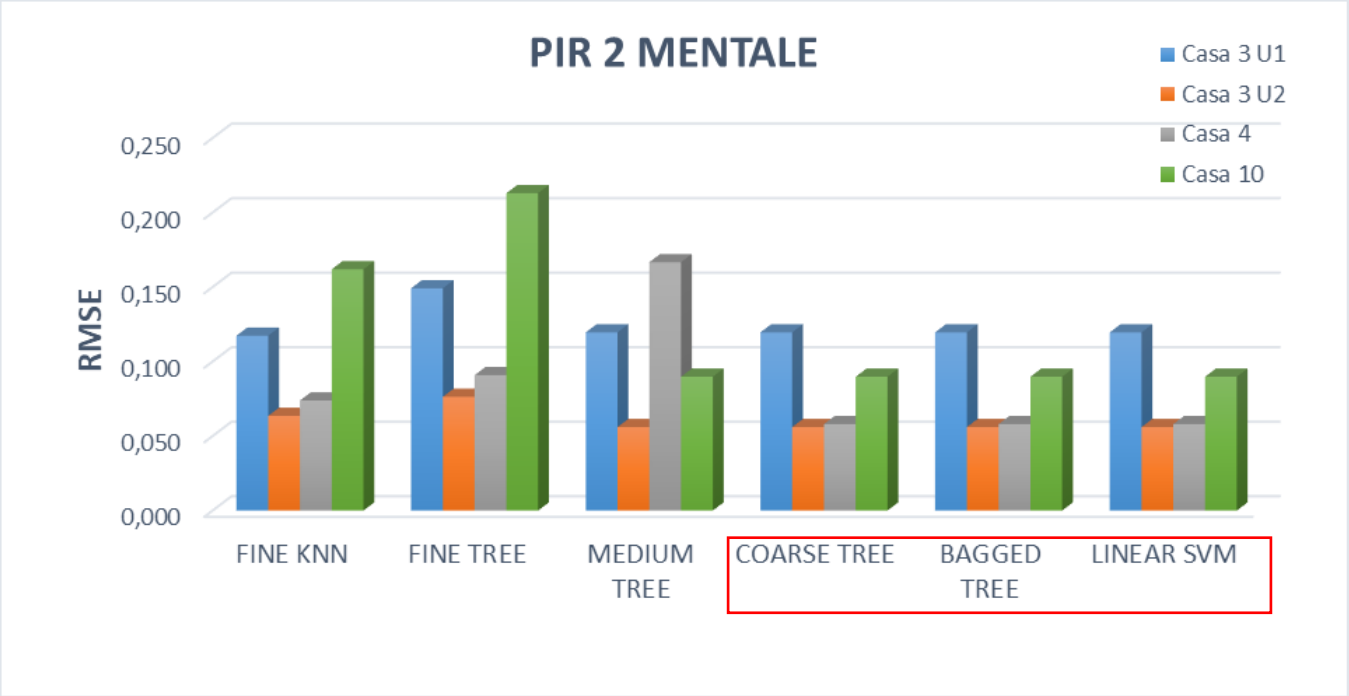

Figura 26: istogramma degli RMSE che esprimono l'accuratezza dei metodi di Machine Learning per predirre il questionario mentale a partire dai dati del PIR 2.

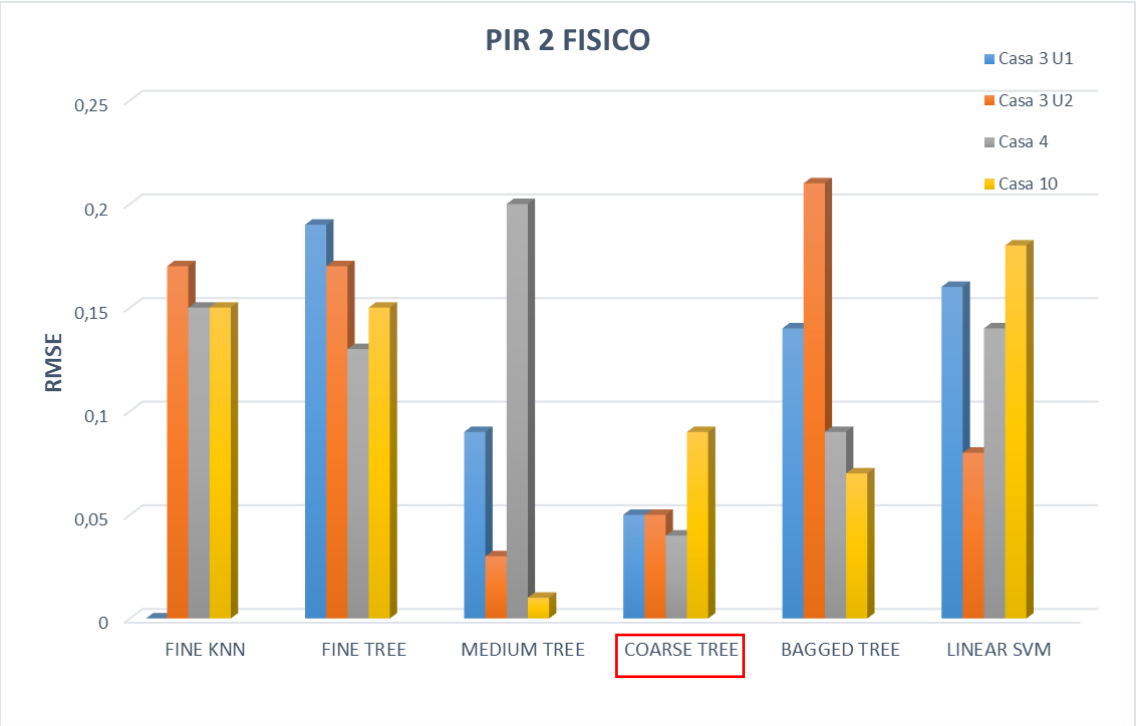

Figura 27: istogramma degli RMSE che esprimono l'accuratezza dei metodi di Machine Learning per predirre il questionario fisico a partire dai dati del PIR 2.

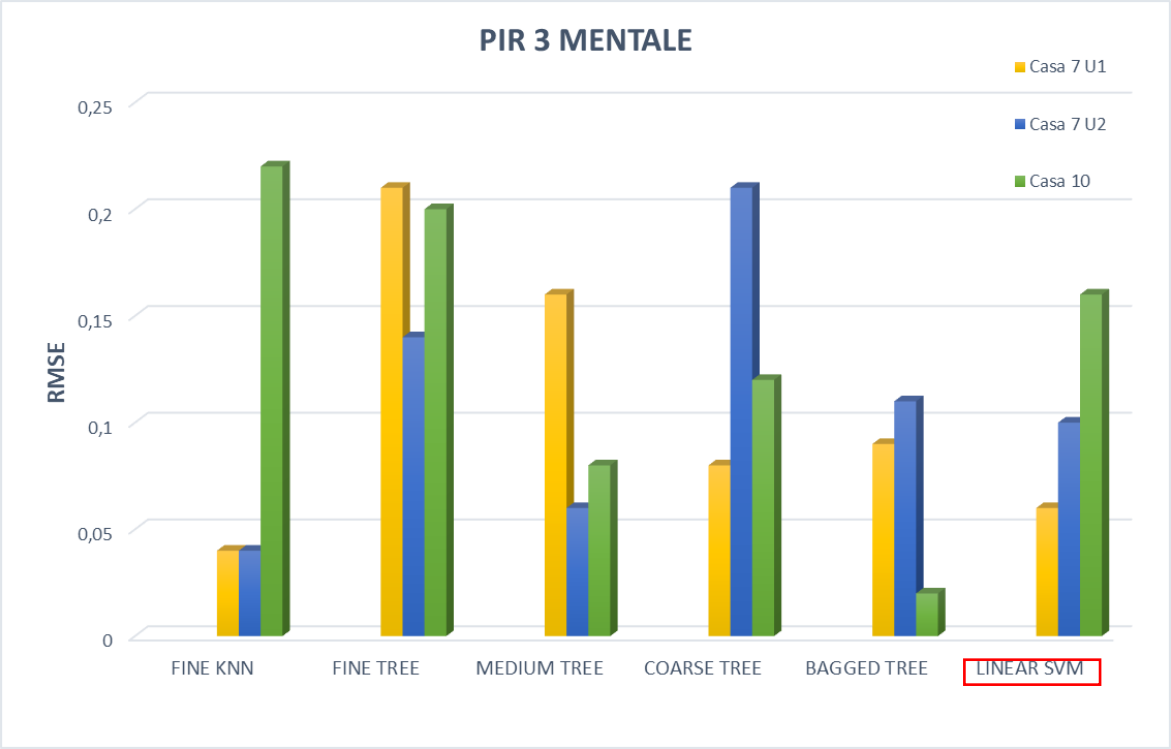

Figura 28: istogramma degli RMSE che esprimono l'accuratezza dei metodi di Machine Learning per predirre il questionario mentale a partire dai dati del PIR 3.

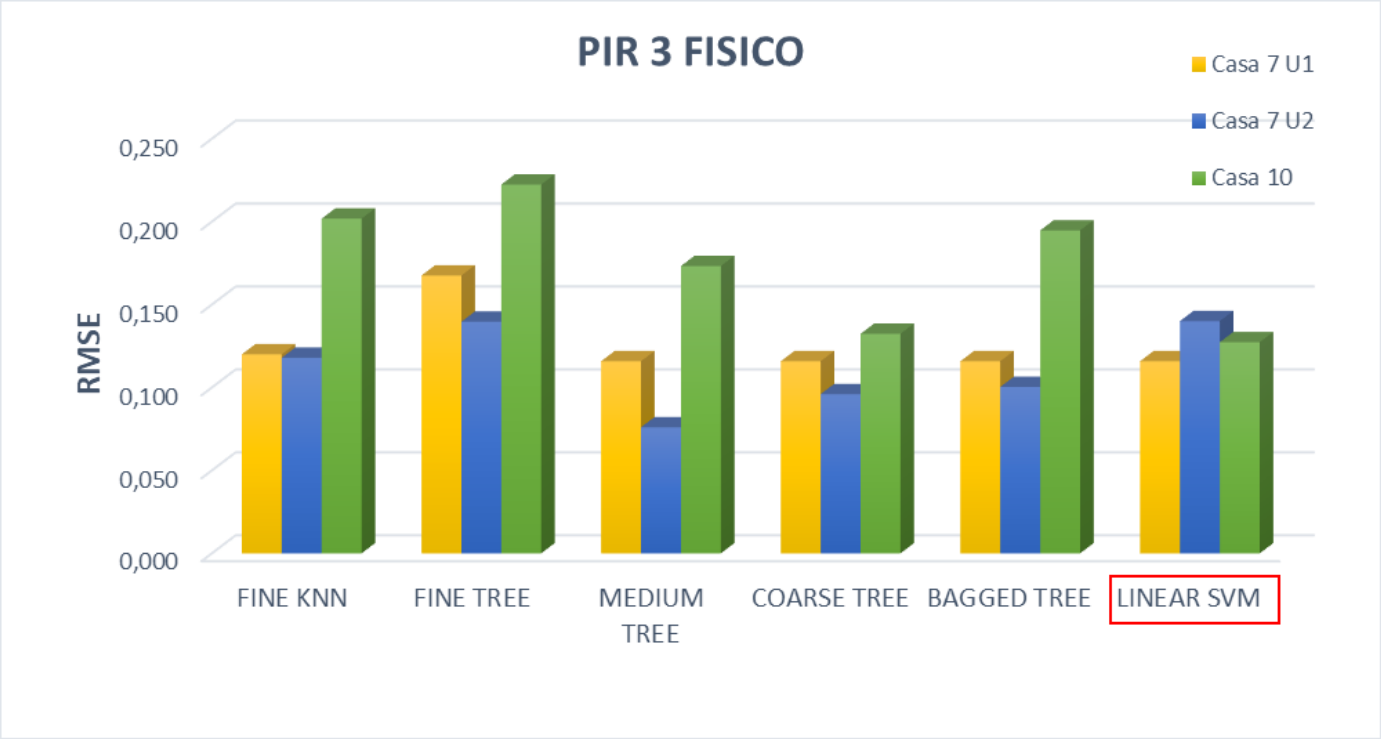

Figura 29: istogramma degli RMSE che esprimono l'accuratezza dei metodi di Machine Learning per predirre il questionario fisico a partire dai dati del PIR 3.

Gli istogrammi sopra riportati sono stati ottenuti andando a graficare gli RSME ricavati mediante le accuratezze ottenute allenando gli algoritmi con i dati relativi all'accensione dei sensori PIR.

Per tutte le stanze prese in considerazione nelle abitazioni si sono registrati, in generale, valori di RMSE minori di 0,25. Considerando gli algoritmi che hanno registrato gli RMSE più bassi evidenziati con un riquadro rosso, si nota che alcuni algoritmi hanno dato risultati migliori, sia per il questionario sullo stato fisico sia per quello sullo stato mentale, in relazione alla parte della casa in cui è stato installato il PIR:

- Il LinearSVM ha restituito gli RSME più bassi quando è stato allenato con le attivazioni del PIR 3 quindi quello relativo al bagno in particolare i valori sono sempre inferiori ai 0,15.
- Il Coarse Tree ha restituito gli RMSE più bassi quando è stato allenato con le attivazioni del PIR 2 ovvero quello relativo alla camera da letto in particolare i valori registrati sono inferiori a 0,12.

Le stesse considerazioni non possono essere fatte per il PIR 1 in cui gli RMSE più bassi sono stati registrati da algoritmi diversi in base al tipo di questionario considerato.

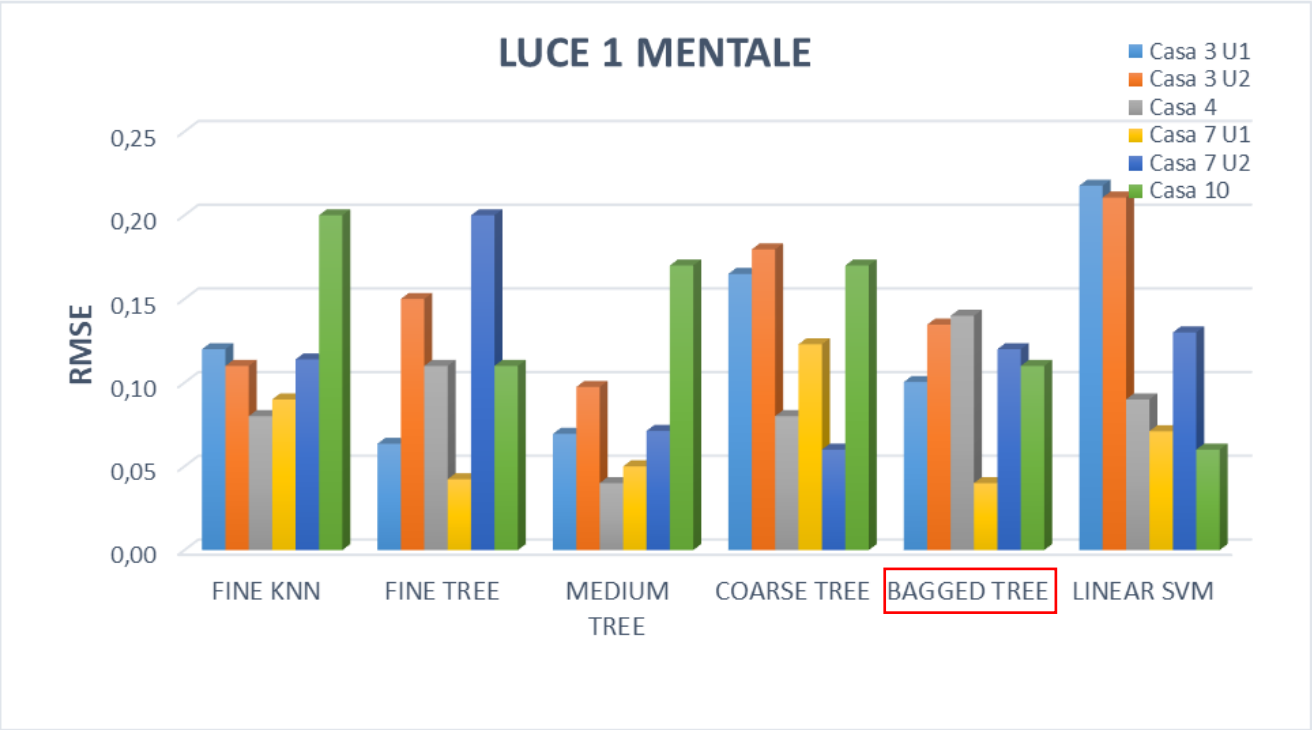

Figura 30: istogramma degli RMSE che esprimono l'accuratezza dei metodi di Machine Learning per predirre il questionario mentale a partire dai dati del sensore luce 1.

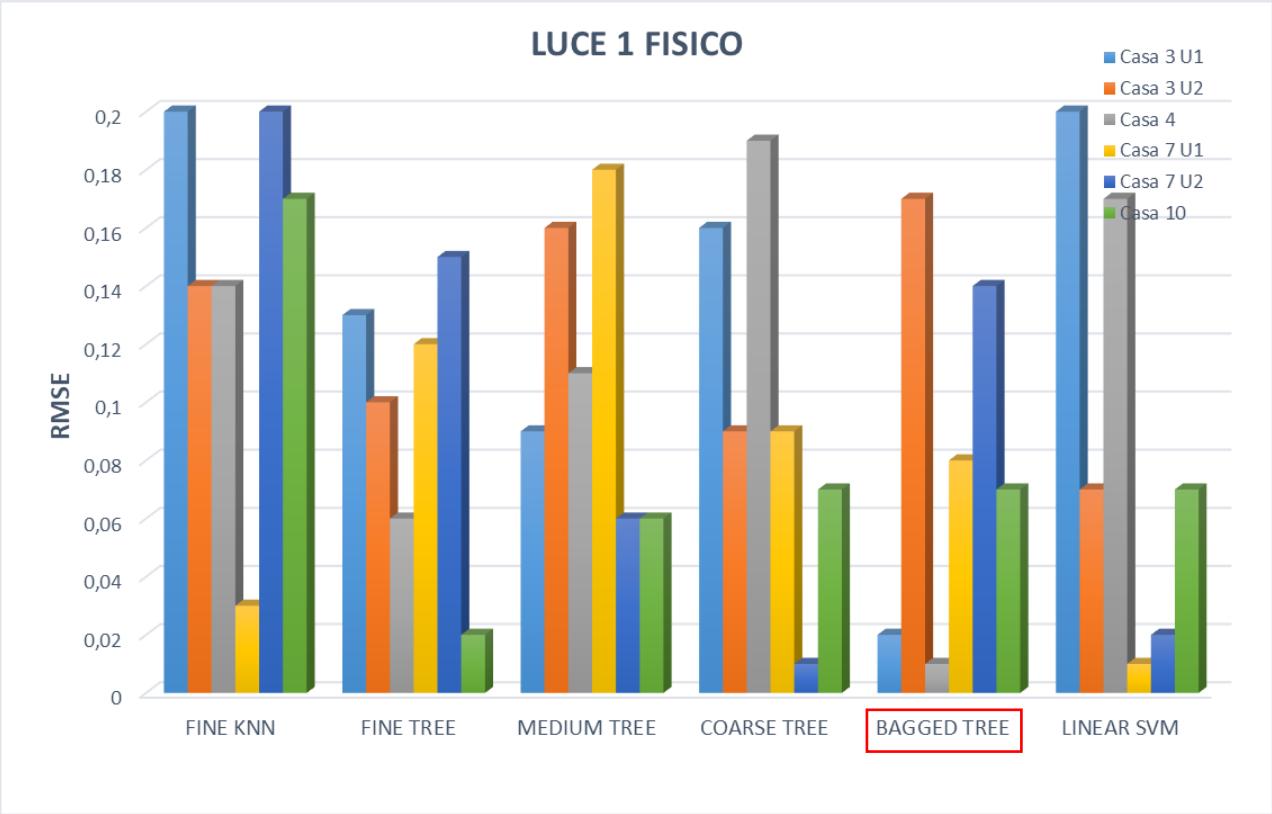

Figura 31: istogramma degli RMSE che esprimono l'accuratezza dei metodi di Machine Learning per predirre il questionario fisico a partire dai dati del sensore luce 1.

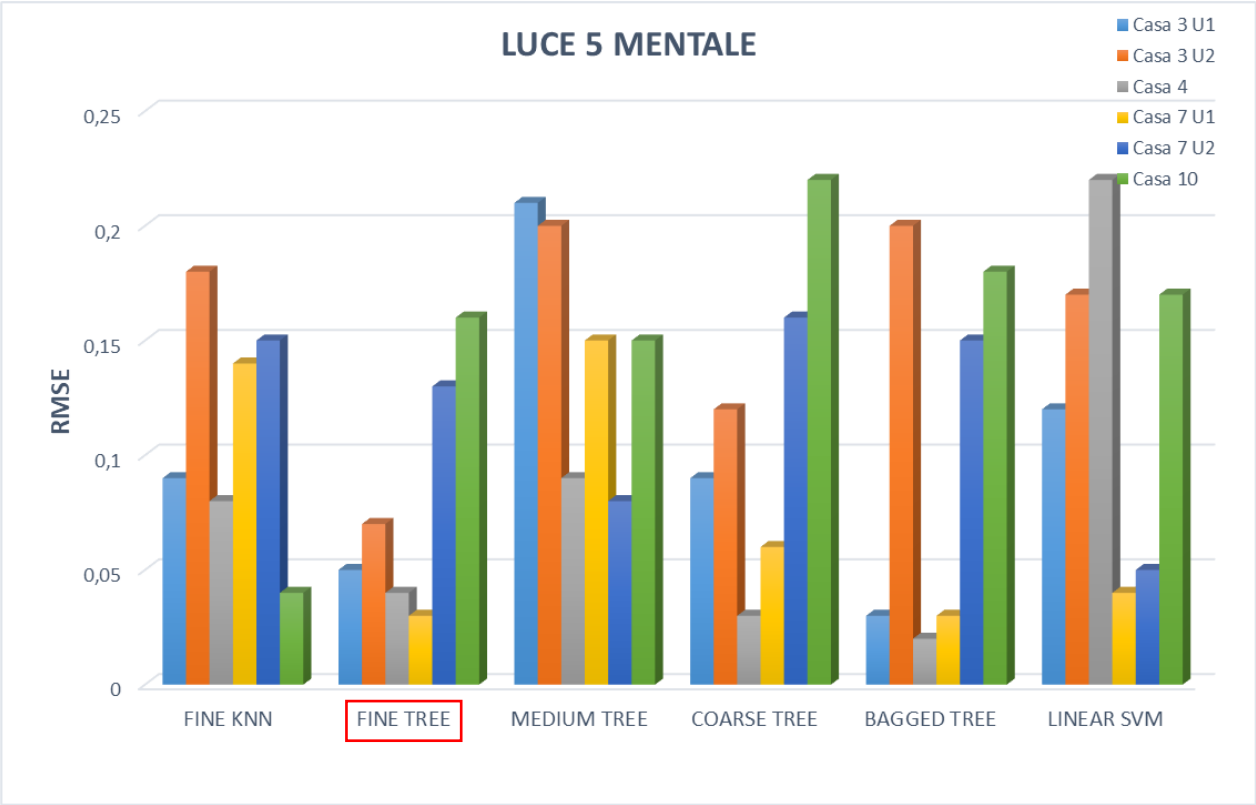

Figura 32: istogramma degli RMSE che esprimono l'accuratezza dei metodi di Machine Learning per predirre il questionario mentale a partire dai dati del sensore luce 5.

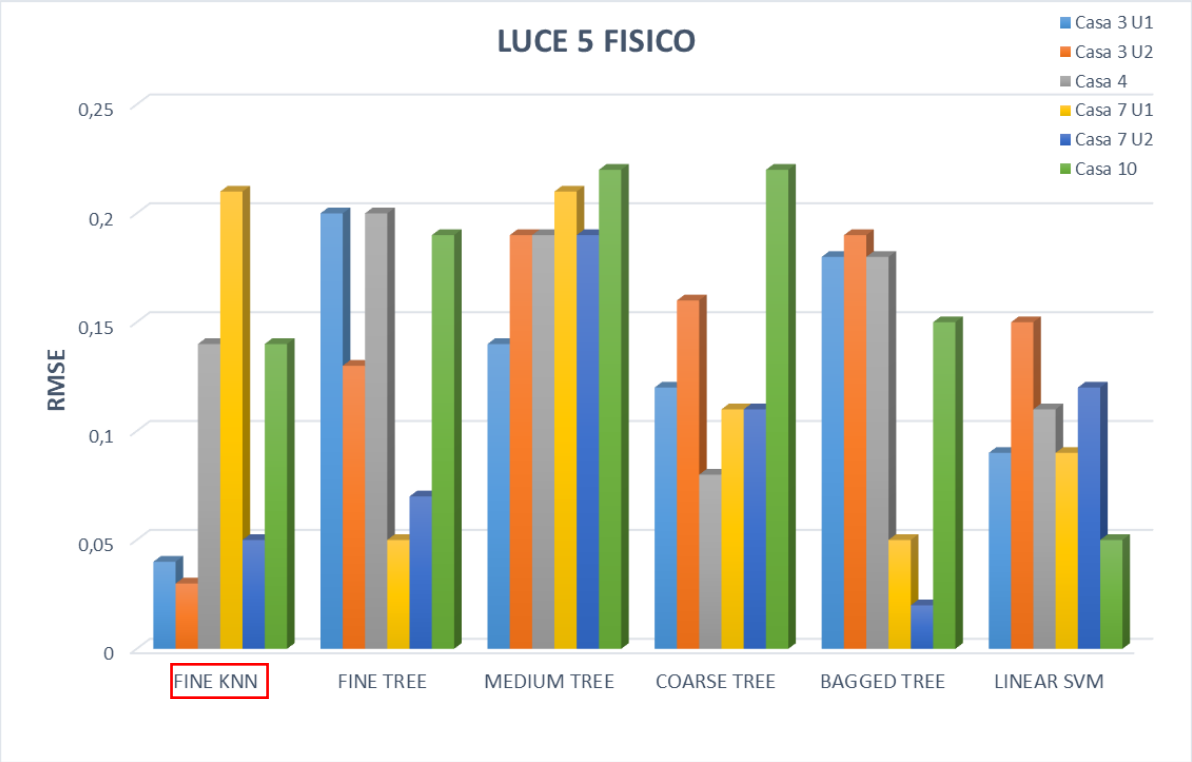

Figura 33: istogramma degli RMSE che esprimono l'accuratezza dei metodi di Machine Learning per predirre il questionario fisico a partire dai dati del sensore luce 5.

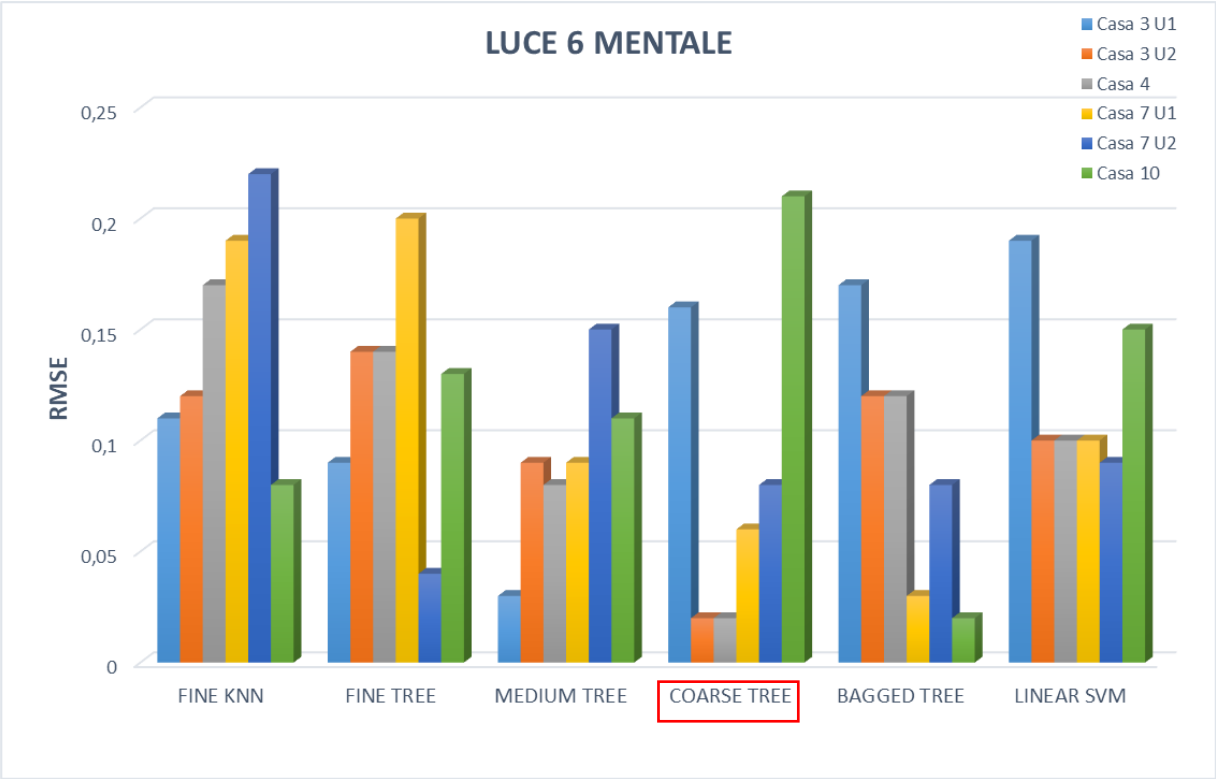

Figura 34: istogramma degli RMSE che esprimono l'accuratezza dei metodi di Machine Learning per predirre il questionario mentale a partire dai dati del sensore luce 6.

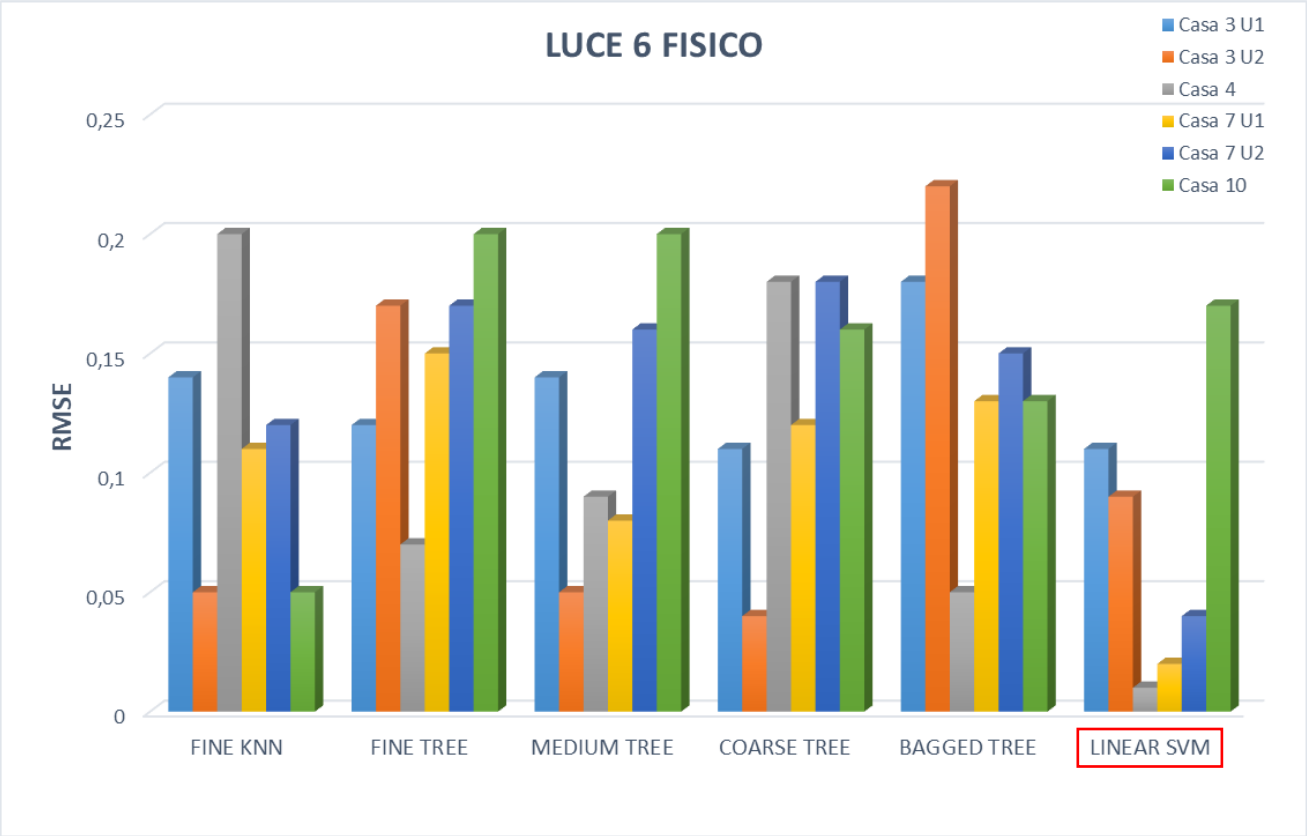

Figura 35: istogramma degli RMSE che esprimono l'accuratezza dei metodi di Machine Learning per predirre il questionario fisico a partire dai dati del sensore luce 6.

Gli istogrammi sopra riportati sono stati ottenuti andando a graficare gli RSME ricavati mediante le accuratezze ottenute allenando gli algoritmi con i dati relativi all'accensione dei sensori luce.

Anche in questo caso sono stati evidenziati gli algoritmi che hanno restituito gli RMSE migliori con un riquadro rosso. Analizzando gli istogrammi delle figure che vanno dalla 30 alla 35, ovvero quelli relativi alle luci, è possibile notare che le validazioni migliori sono state ottenute mediante Linearsvm e Bugged Tree. L'abitazione numero quattro ha infatti ottenuto RMSE molto piccoli (circa 0,05) sia nella predizione del questionario fisico sia mentale.

Dando uno sguardo generale ai dati relativi sia ai sensori PIR che ai sensori luce, si registrano due casi in cui gli algoritmi che hanno dato i risultati migliori in tutte le case all'interno della stessa stanza sono gli stessi per la stessa tipologia di questionario: il primo è il Bugged Tree che ha registrato per il questionario fisico i valori migliori in salotto ed il secondo è il LinearSVM che ha registrato gli RMSE del questionario fisico più bassi nel bagno.

6. Conclusioni

Questo studio condotto nell'ambito del progetto Health@Home ha avuto come fine quello di dimostrare che attraverso una rete integrata di sensori domotici quali, sensori PIR e sensori luci, si può predirre il benessere degli utenti.

La sperimentazione del progetto H@H nelle 8 abitazioni del comune di Oderzo ha permesso di dimostrare che attraverso l'istallazione di sensori domotici, la raccolta dei dati, la loro analisi e i questionari forniti si può analizzare il comportamento di persone anziane all'interno delle loro abitazioni per identificare i servizi da fornire a quest'ultimi.

Le tecniche utilizzate per l'analisi dei dati raccolti hanno permesso di gestire in maniera ottimale la grande mole di quest'ultimi, cosa che sarebbe stata impossibile senza l'utilizzo di software come Matlab. Proprio grazie a questo è stato possibile l'utilizzo del Machine Learning, che ha avuto un ruolo fondamentale per lo studio del comportamento dei soggetti. L'utilizzo di tali tecniche, anche in campo domotico, sta avendo sempre più successo in quanto permettono di identificare, analizzare e comprendere possibili problemi e relative soluzioni sia per quanto riguarda l'assistenza sanitaria rivolta agli utenti più anziani, ai disabili o ad individui con problematiche di qualsiasi genere, sia per gli aspetti domotici in toto.

Lo scopo di questa tesi è trovare una metodologia che possa correlare i dati domotici al questionario che esprime il benessere mentale e fisico di una persona. Per questo scopo sono state dapprima cercate delle possibili correlazioni tra le attivazioni dei sensori PIR e i dati provenienti dal questionario, tuttavia questo percorso si è rivelato inconcludente a causa dei bassi valori di RHO_S e RHO_P. Alla luce di ciò sono state infine utilizzate tecniche di Machine Learning in seguito validate tramite cross-validation e hold-out validation. La prima tecnica di valutazione ha portato dei buoni risultati soprattutto predicendo, grazie agli alberi decisionali, il questionario mentale. Ad esempio, si è riusciti a predirre lo stato mentale dell'utente due di casa 3 con una precisione del 90%, non si possono fare le stesse conclusioni per il questionario fisico. Data la discrepanza tra le accuratezze del questionario mentale e fisico, le validazioni sono state fatte mediante hold-out validation. Questo tipo di analisi ha portato risultati molto soddisfacenti come ad esempio l'RMSE registrato tra il valore predetto dal Medium Tree tramite le attivazioni della luce 6 e il valore vero del questionario fisico dell'utente 2 dell'abitazione numero 3 pari a 0,05. In generale questo metodo di validazione ha registrato RMSE bassi tutti inferiori a 0,25.

La procedura presentata in questa tesi può essere un punto di partenza per future analisi in campo domotico ma non solo, in quanto adattabile a qualunque tipo di segnale. L'avanzare della tecnologia porterà questo tipo di studio a livelli ancora più elevati, dando la possibilità di fornire soluzioni sempre più performanti dal punto di vista assistenziale e quindi di migliorare la permanenza degli individui all'interno di abitazioni ed edifici in generale.

Bibliografia

[1] Bennett, J., Rokas, O., & Chen, L. (2017). Healthcare in the Smart Home: A Study of Past, Present andFuture. Sustainability, 9(5), 840.

[2] Administration on Aging (AoA), U.S. Department of Health and Human Services. (2008, February 6). A Statistical Profile of Older Americans Aged 65+

[3] Galluzzo, L., Gandin, C., Ghirini, S., & Scafato, E. (2018). L'invecchiamento della popolazione:opportunità o sfida. archivio, 2017, 2016.

[4] Madhusudhanan, D., Balaji, R., & Manikandan, M. (2014, July). Casattenta, Sofia and Tangerine &IT4CAD: Practical experiences. In Current Trends in Engineering and Technology (ICCTET), 2014 2ndInternational Conference on (pp. 67-73). IEEE.

[5] Wilson, C., Hargreaves, T., & Hauxwell-Baldwin, R. (2017). Benefits and risks of smart hometechnologies. Energy Policy, 103, 72-83.

[6] Hargreaves, T., Wilson, C., & Hauxwell-Baldwin, R. (2018). Learning to live in a smart home. BuildingResearch & Information, 46(1), 127-139.

[7] Unsupervised Machine Learning for DevelopingPersonalised Behaviour Models Using Activity DataLaura Fiorini1,*, Filippo Cavallo1, Paolo Dario1, Alexandra Eavis2and Praminda Caleb-Solly31The BioRobotics Institute, Scuola Superiore Sant'Anna, Pontedera, Pisa 56025, Italy;filippo.cavallo@santannapisa.it (F.C.); paolo.dario@santannapisa.it (P.D.)2Alcove Limited, 44 Westbridge Road, London SW11 3PW, UK; alexandra.eavis@youralcove.com3Bristol Robotics Laboratory, University of West England, Bristol BS16

[8] A Plug and Play IoT Wi-Fi Smart Home System for Human Monitoring ; Marco Bassoli , Valentina Bianchi * and Ilaria De Munar

[9] Majumder, S., Aghayi, E., Noferesti, M., Memarzadeh-Tehran, H., Mondal, T., Pang, Z., & Deen, M. J.(2017). Smart Homes for Elderly Healthcare—Recent Advances and Research Challenges. Sensors, 17(11),2496.

[10] Wu, T., Wu, F., Redoute, J. M., & Yuce, M. R. (2017). An autonomous wireless body area networkimplementation towards IoT connected healthcare applications. IEEE Access, 5, 11413-11422.

[11] Benetazzo, F., Ferracuti, F., Freddi, A., Giantomassi, A., Iarlori, S., Longhi, S., ... & Ortenzi, D. (2015).AAL technologies for independent life of elderly people. In Ambient Assisted Living(pp. 329-343). Springer,Cham.

[12] Casaccia, S., Pietroni, F., Scalise, L., Revel, G. M., Monteriù, A., Prist, M. R., ... & Longhi, S. (2018,June). Health@ Home: pilot cases and preliminary results: Integrated residential sensor network to promotethe active aging of real users. In 2018 IEEE International Symposium on Medical Measurements andApplications (MeMeA) (pp. 1-6). IEEE.

[13] Smart monitoring of user and home environment: the Health@Home acquisition framework. Foritaal 2017, Genova.Sara Casaccia, Filippo Pietroni, Lorenzo Scalise, Gian Marco Revel, Andrea Monteriù, Mario Rosario Prist, Emanuele Frontoni, Sauro Longhi.

[14] Health@Home: pilot cases and preliminary results. MEMEA 2018, Rome. Monteriù, A., Prist, M., Frontoni, E., Longhi, S., Pietroni, F., Casaccia, S., ... & Pescosolido, L. (2018). A Smart Sensing Architecture for Domestic Monitoring: Methodological Approach and Experimental Validation. Sensors, 18(7), 2310.

[15] Jaeseok, Y.; Sang-Shin, L., (2014, May). Human mouvment detection and idengification using Pyroelectric Infrared Sensors.

[16] Scalise, L.; Pietroni, F.; Casaccia, S.; Revel, G.M.; Monteriù, A.; Prist, M.; Longhi, S.; Pescosolido, L. Implementation of an "at-home" e-health system using heterogeneous devices. In Proceedings of the IEEE International Smart Cities Conference, Trento, Italy, 12–15 September 2016; pp. 1–4.

[17] Varalda, A., (2017, October). Domotica assistenziale-la smart home per chi ha più bisogno.

[18] Scalise, L.; Pietroni, F.; Casaccia, S.; Revel, G.M.; Monteriù, A.; Prist, M.; Longhi, S.; Frontoni, E. Healt@Home: pilot cases and preliminary results. (2017).

[19] Andrea Monteriù, Mario Rosario Prist, Emanuele Frontoni, Sauro Longhi, Filippo Pietroni, Sara Casaccia, Lorenzo Scalise, Annalisa Cenci, Luca Romeo, Riccardo Berta, Loreto Pescosolido, Gianni Orlandi and Gian Marco Revel. A Smart Sensing Architecture for Domestic Monitoring: Methodological Approach and Experimental Validation. (2018, july).

[20] Spearman C., 1904. The proof and measurement of association between two things. American Journal of Psychology 15: 72–101.

[21] https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statisticalguide.php

[22] https://www.ai4business.it/intelligenza-artificiale/machine-learning/machine-learning-cosae-applicazioni/

[23] https://www.ai4business.it/intelligenza-artificiale/machine-learning/machine-learning-cosae-applicazioni/

[24] Cos'è il Machine Learning (Reti Neurali e A.I.) https://www.youtube.com/watch?v=X_ZUyvAA9Qw

[25] https://jacopokahl.com/i-tre-principali-tipi-di-machine-learning-apprendimentosupervisionato-non-supervisionato-e-per-rinforzo/

[26] https://www.spaghettiml.com/2017/10/03/decision-tree-o-alberi-decisionali/

[27] https://lorenzogovoni.com/support-vector-machine/

[28] https://stackabuse.com/k-nearest-neighbors-algorithm-in-python-and-scikit-learn/

[29]Confusion Matrix-based Feature Selection Sofia Visa Computer Science Department College of Wooster Wooster, OH svisa@wooster.edu Brian Ramsay Sentry Data Systems, Inc. Fort Lauderdale, FL brian.ramsay@gmail.com Anca Ralescu Computer Science Department University of Cincinnati Cincinnati, OH anca.ralescu@uc.edu Esther van der Knaap Ohio Agricultural Research and Development Center The Ohio State University Wooster, OH vanderknaap.1@osu.edu

[30] http://www.andreaminini.com/ai/machine-learning/overfitting

[31] Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification Sanjay Yadav ; Sanyam Shukla.

[32] https://www.datavedas.com/holdout-cross-validation/